DOCUMENT RESUME

ED 476 816 PS 031 310

| | |
|---|---|
| TITLE | Assessing the State of State Assessments: Perspectives on Assessing Young Children. Special Report. |
| INSTITUTION | SERVE: SouthEastern Regional Vision for Education, Greensboro, NC. |
| SPONS AGENCY | National Inst. of Education (ED), Washington, DC. |
| PUB DATE | 2003-00-00 |
| NOTE | 78p. |
| CONTRACT | ED-01-CO-0015 |
| AVAILABLE FROM | SERVE, P.O. Box 5367, Greensboro, NC 27435. Tel: 800-755-3277 (Toll Free); Tel: 336-315-7400; Fax: 336-315-7457; Web site: http://www.serve.org. |
| PUB TYPE | Collected Works - Proceedings (021) -- Reports - Evaluative (142) |
| EDRS PRICE | EDRS Price MF01/PC04 Plus Postage. |
| DESCRIPTORS | Early Childhood Education ; Educational Policy; Evaluation Criteria; *Evaluation Methods; Kindergarten; Learning Readiness; Measures (Individuals); *Preschool Children; *School Readiness; *Student Evaluation |
| IDENTIFIERS | Project Head Start |

ABSTRACT

A number of preschool and kindergarten assessment systems are being put into place across the nation, with a variety of purposes and collection methods. The "Assessing the State of State Assessments" symposium was convened to provide an opportunity for persons working most closely with state assessment systems to identify common challenges and share ideas. This special report presents a compilation of perspectives on assessment issues discussed at the symposium. Chapter 1, "Assessing Young Children: What Policymakers Need To Know and Do" (Sharon Lynn Kagan, Catherine Scott-Little, and Richard M. Clifford), reviews basic principles that should guide early childhood assessment policies and outlines critical policy issues related to assessment systems. Chapter 2, "A Risk Management Approach to Readiness Assessment: Lessons from Florida" (Susan Muenchow), defines several readiness assessment terms and presents four potential benefits of readiness assessment systems. Chapter 3, "Assessing School Readiness: System Design Framework and Issues" (Gary T. Henry), argues that a key design element of an assessment system is discerning the purpose for which the assessments are being conducted, and presents important issues that should be taken into consideration when making assessment design decisions. Chapter 4, "Issues in Implementing a State Preschool Program Evaluation in Michigan" (Lawrence J. Schweinhart), addresses a variety of design and implementation issues encountered by the High/Scope Educational Research Foundation's evaluation of the Michigan School Readiness Program. The final chapter, "Instrumentation for State Readiness Assessment: Issues in Measuring Children's Early Development and Learning" (John M. Love), examines technical issues related to assessment, providing a system-level review of the elements of readiness, along with current political and education factors that affect readiness assessment systems. The chapter includes criteria for evaluating both an individual measure as well as a set of measures used in an assessment system. Two attachments include "Readiness" dimensions identified by the Goal One Technical Planning Group of the National Educational Goals Panel, and discussion of how Head Start Performance Measures are aligned with the Goal One Dimensions. The document concludes with [illegible] challenges and

Reproductions supplied by EDRS are the best that can be made
from the original document.

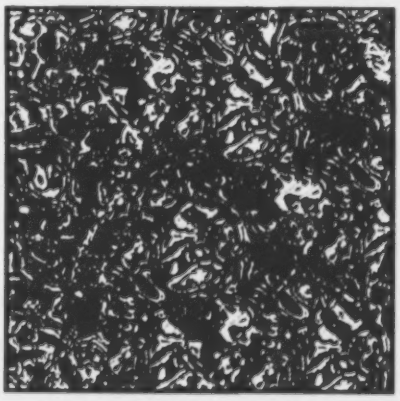
next steps of statewide school readiness assessments (Martha Zaslow and Tamara Halle). (HTH)

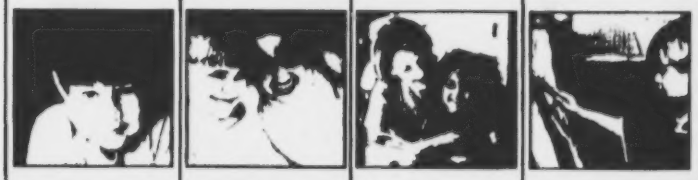
Reproductions supplied by EDRS are the best that can be made
from the original document.

# Assessing the State of State Assessments:

## Perspectives on Assessing Young Children

ED 476 816





PS 031310

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

SERVE™
*Improving Learning through Research & Development*

# *Assessing the State of State Assessments:*

## Perspectives on Assessing Young Children

*A special report on wide-scale early childhood assessment systems*

*Edited for the Regional Educational Laboratory at SERVE by:*

**Catherine Scott-Little**
The Regional Educational Laboratory at SERVE

**Sharon Lynn Kagan**
Teachers College, Columbia University

**Richard M. Clifford**
The National Center for Early Development and Learning

# *Credits*

***Produced by***

The REGIONAL EDUCATIONAL LABORATORY at SERVE℠

Associated with the School of Education,
University of North Carolina at Greensboro

***First Printing 2003***

***Edited by***

**Donna Nalley**, Publications Director, SERVE
**Karen DeMeester**, Senior Program Specialist, SERVE
**Christy Casbon**, Assistant Program Specialist, SERVE

***Designed by***

**Jeremy McClellan**, Graphic Designer, SERVE

The content of this publication does not necessarily reflect the views or policies of the National Institute for Education Sciences, U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

This document was produced with funding from the National Institute for Education Sciences, U.S. Department of Education, under contract no. ED-01-CO-0015.

# *Acknowledgments*

This document is a collection of papers that reflect discussions that took place during a symposium on readiness assessment systems and the authors' collective experience working on issues related to wide-scale early childhood assessment systems. Special thanks go to the two organizations that funded the symposium: The A. L. Mailman Family Foundation and the National Institute for Education Sciences of the U.S. Department of Education.

We would also like to thank the following contributing authors for their thoughtful work that is reflected in this report:

**Tamara Halle,** Child Trends
**Gary T. Henry,** Andrew Young School of Policy Studies, Georgia State University
**John M. Love,** Mathematica Policy Research
**Susan Muenchow,** American Institutes for Research
**Lawrence J. Schweinhart,** High/Scope Educational Research Foundation
**Martha Zaslow,** Child Trends

The following reviewers for this document provided helpful suggestions for which we are grateful:

**Susan Andersen,** Iowa Department of Education
**Glyn Brown,** The Regional Educational Laboratory at SERVE
**Donna Bryant,** Frank Porter Graham Child Development Institute
**Beth Garriss Hardy,** The Regional Educational Laboratory at SERVE
**Marilou Hyson,** National Association for the Education of Young Children
**Samuel Meisels,** The Erikson Institute

We also would like to thank the following support staff who were instrumental in making arrangements for the symposium:

Louise Scrivani, Rosemarie Andrews, Michael Spencer, and Amanda Payne

**Finally, we would like to thank the symposium participants listed below for their ongoing work on behalf of children and families and their willingness to share their experiences and insights during the symposium.**

**Cindy Bagwell**
North Carolina Department of Public Instruction

**Betty Bardige**
A. L. Mailman Family Foundation

**Glyn Brown**
The Regional Educational Laboratory at SERVE

**Donna Bryant**
Frank Porter Graham Child Development Institute

**Lindy Buch**
Michigan Department of Education

**Richard M. Clifford**
Frank Porter Graham Child Development Institute

**Dawn Denno**
Ohio Department of Education

**Lucy Dwight**
National Conference of State Legislatures

**Laurie Ford**
University of British Columbia

**Theresa Garcia**
California Office of Secretary for Education

**Rolf Grafwallner**
Maryland State Department of Education

**Sarah Greene**
National Head Start Association

**Tamara Halle**
Child Trends

**Doris Hallford**
Missouri Division of Family Services

**Sharon Hawley**
California Department of Education

**Jane Henderson**
California Children and Families Commission

**Gary Henry**
Andrew Young School of Policy Studies,
Georgia State University

**A. Baron Holmes**
South Carolina Budget & Control Board

**Marilou Hyson**
National Association for the Education
of Young Children

**Sharon Lynn Kagan**
Teachers College, Columbia University

**Katherine Kamiya**
Florida Partnership for School Readiness

**Naomi Karp**
Institute of Education Sciences

**John Love**
Mathematica Policy Research, Inc.

**Kelly Maxwell**
Frank Porter Graham Child Development Institute

**Kerry Mazzoni**
California Office of Secretary for Education

**Kathleen Dufford-Melendez**
The Regional Educational Laboratory at SERVE

**Sandra Miller**
Ohio Office of Early Childhood Education

**Linda Mims**
South Carolina Department of Education

**Marsha Moore**
Georgia Office of School Readiness

**Susan Muenchow**
American Institutes for Research

**John R. Sanders**
The Regional Educational Laboratory at SERVE

**Sharon Ford Schattgen**
Missouri Department of Elementary and
Secondary Education

**Larry Schweinhart**
High/Scope Educational Research Foundation

**Catherine Scott-Little**
The Regional Educational Laboratory at SERVE

**Pam Shapiro**
Georgia Office of School Readiness

**Sandra Skolnik**
Maryland Committee for Children

**Helene Stebbins**
National Governors Association

**Kathy R. Thornburg**
Child Development Laboratory, University of Missouri

**Marce Verzaro-O'Brien**
Region IV Head Start Quality Improvement Centers

**Jerry Walker**
Ohio Department of Education

**Catherine Woodall**
North Carolina Partnership for Children

**Marty Zaslow**
Child Trends

# *Table of Contents*


| | Page |
|---|---|
| *Acknowledgments* | iii |
| *Preface* | 1 |
| Assessing Young Children: What Policymakers Need to Know and Do<br>Sharon Lynn Kagan, Teachers College, Columbia University<br>Catherine Scott-Little, The Regional Educational Laboratory at SERVE<br>Richard M. Clifford, The National Center for Early Development and Learning | 5 |
| A Risk Management Approach to Readiness Assessment: Lessons from Florida<br>Susan Muenchow, American Institutes for Research | 13 |
| Assessing School Readiness: System Design Framework and Issues<br>Gary T. Henry, Andrew Young School of Policy Studies, Georgia State University | 25 |
| Issues in Implementing a State Preschool Program Evaluation in Michigan<br>Lawrence J. Schweinhart, High/Scope Educational Research Foundation | 37 |
| Instrumentation for State Readiness Assessment: Issues in Measuring Children's Early Development and Learning<br>John M. Love, Mathematica Policy Research | 43 |
| Attachment One:<br>"Readiness" Dimensions Identified by the Goal One Technical Planning Group of the National Education Goals Panel | 51 |
| Attachment Two:<br>How Head Start Performance Measures are Aligned with the Goal One Dimensions | 53 |
| Statewide School Readiness Assessments: Challenges and Next Steps<br>Martha Zaslow and Tamara Halle, Child Trends | 57 |
| About the Editors | 67 |
| About the Authors | 69 |
| About the National Center for Children and Families: Advancing Policy, Education, and Development | 71 |
| About the National Center for Early Development and Learning | 75 |
| About SERVE | 77 |

# *Preface*

Assessment of children around the age of kindergarten entry has become a "hot" topic for educators, researchers, and policymakers. Pressure for wide-scale assessments to collect data from large numbers of preschool and kindergarten children is mounting at both the federal and state level. On the federal level, Head Start and Even Start regulations require that assessment data be collected from very young children. State-level assessment systems are also being developed, and with these efforts come challenges with regard to implementing assessment systems. For example, a national survey (Saluja, Scott-Little, & Clifford, 2000) found 13 states had established statewide screening or assessment programs for children entering kindergarten in the fall of 1999. Five additional states required statewide screenings or assessments but allowed local districts to decide how to conduct the assessments. An additional 16 states had initiatives in place at the time of the survey to develop recommendations for how children should be assessed. Clearly, there is an increasing tendency for wide-scale assessments to be conducted with very young children.

Developing and implementing such assessment systems, at any level in any program, is not an easy matter. With the increasing pressure to collect data on the proficiency of large numbers of young children in particular areas of development has come a corresponding increase in concerns about the purpose of the assessments, the nature of assessment processes, and the implications for how the data are being used. Why are assessments being conducted, and are states clear on the purposes of such assessment? Do they distinguish between, for example, assessment to improve instruction and assessment to make high-stakes decisions about children or programs? How can assessment data be collected on a wide scale in a manner that is technically sound and beneficial for both the children and the stakeholders interested in using the results? And how are the data being used? Is the use matching the original intentions, or are the instruments designed for one purpose being used for another?

These questions plague the minds of policymakers, assessment specialists, early childhood educators, and even parents. Complicating the issue even further is the well-documented fact that assessment of young children is difficult and requires specialized techniques. Young children's inability to read, the episodic nature of their learning, and their stress in unfamiliar settings with unfamiliar people all contribute to the special challenges facing those concerned about assessment of young children. Finally, many argue that there are a very limited number of suitable assessments for effectively measuring the domains that are of great importance to the developing young child; emotional development and approaches toward learning are noteworthy in this respect.

In trying to develop assessment systems and practices that are well-constructed and fair, states have struggled. At best, there is potential for these assessment systems to produce credible and useful information. At worst, they can produce unintended negative consequences for children and programs. In designing such systems, states have struggled to at minimum "do

no harm." There are no roadmaps for the development of early childhood assessment systems. Such wide-scale assessment systems are a new undertaking fraught with challenges and deplete of models and experiences to draw from. States are blazing new trails as they attempt to develop assessment systems that are sensitive to the information needs of policymakers and programs and, at the same time, the needs of children.

Despite the challenges, states are developing wide-scale assessment systems to collect data on children's skills and characteristics around the age of kindergarten entry. A number of assessment systems are being put into place across the nation, with a variety of purposes and data-collection methodologies. In order to learn from states that are developing such systems and to identify possible next steps to support efforts to develop sound assessment systems, representatives from nine states with experience in establishing such systems were invited to participate in a symposium on assessment systems. The symposium, entitled *Assessing the State of State Assessments*, was designed to provide an opportunity for persons working most closely with state assessment systems to identify common challenges and share ideas. With funding from the A. L. Mailman Family Foundation and the U.S. Department of Education, Office of Education Research and Improvement (OERI), Dr. Sharon Lynn Kagan of Teachers College, Columbia University; Dr. Richard M. Clifford from the National Center for Early Development and Learning; and Dr. Catherine Scott-Little from the Regional Educational Laboratory at SERVE hosted the symposium on December 12–14, 2001. Teams from California, Georgia, Florida, Maryland, Michigan, Missouri, North Carolina, Ohio, and South Carolina came together with researchers and representatives from national organizations such as the National Association for the Education of Young Children (NAEYC), the National Governor's Association (NGA), and the National Conference of State Legislatures (NCSL) to discuss issues associated with establishing wide-scale assessment systems for young children.

Prior to the symposium, several focus groups were held to determine the most critical issues that states face as they develop such systems. As a result of the focus groups, four issue categories emerged:

- Design issues
- Instrumentation issues
- Implementation issues
- Data utilization issues

During the symposium, participants had the opportunity to share their challenges within each of the categories and to learn from each other. The result was a fruitful discussion of common challenges and possible solutions for states that are putting wide-scale early childhood assessment systems into place.

This document presents a compilation of perspectives on assessment issues discussed during the *Assessing the State of State Assessment Systems* symposium. Four papers were commissioned prior to the symposium to provide a framework for symposium discussions, one to address each of the four issue categories identified above. Two papers, one on next steps for the early childhood assessment field and one on implications for policymakers, were prepared after the symposium to synthesize the issues discussed. The result is this special report on wide-scale early childhood assessment systems that is designed to address the four critical areas—design issues, instrumentation issues, implementation issues, and data utilization issues—from a variety of perspectives. This document does not purport to have covered all the issues related to wide-scale early childhood assessment systems. Rather, it is a compilation of issues that were most salient for the group of persons attending the symposium. Likewise, this is not a document that provides "solutions" or "answers" to many of the issues that plague efforts

to develop wide-scale assessment systems for young children. Rather, it is an edited volume, with diverse perspectives represented. Indeed, readers will find overlapping as well as contradictory perspectives on various issues across the chapters. The presence of contradicting views on early childhood assessment issues is an indicator of the complexity of the issues being addressed. We hope that we have been successful in raising issues that are significant for the field in order to stimulate further discussion.

Readers will find that chapters in this volume include both broad issues related to purposes of assessment systems and policies, as well as narrower, more technical issues such as instrumentation. The first chapter, *Assessing Young Children: What Policymakers Need to Know and Do* by Sharon Lynn Kagan, Catherine Scott-Little, and Richard M. Clifford, reviews basic principles that should guide early childhood assessment policies and outlines critical policy issues related to assessment systems. Given the complexities of early childhood assessment and the increasing need for credible and reliable information about the skills and characteristics of very young children, policymakers face the dilemma of developing policies that can both produce the data needed and protect the well being of children and early childhood programs. This chapter offers critical issues for consideration by policymakers as they promulgate assessment systems that are appropriate and effective.

The next chapter, *A Risk Management Approach to Readiness Assessment: Lessons from Florida* by Susan Muenchow, defines several readiness assessment terms and presents four potential benefits of readiness assessment systems. Drawing from experiences in designing and implementing a readiness assessment system in Florida, Muenchow outlines several potential issues that can lead to unintended negative consequences for such assessment systems and then suggests strategies for minimizing the potential risk of unintended negative consequences. Included among the strategies are a set of principles to guide the development and implementation of Florida's readiness assessment system.

Following these two chapters that address broader issues, the discussion turns to more technical issues associated with the design and implementation of early childhood assessment systems. In *Assessing School Readiness: System Design Framework and Issues*, Gary Henry argues that a key design element of an assessment system is discerning the purpose for which the assessments are being conducted. The assessment system design should flow from the purpose. Henry discusses one purpose—informing the public and policymakers about the adequacy of societal investments in children's earliest years—in detail. Important issues that should be taken into consideration when decisions are made about the design of an assessment system are also presented.

In the following chapter titled *Issues in Implementing a State Preschool Program Evaluation in Michigan*, Lawrence Schweinhart addresses a variety of design and implementation issues encountered by the High/Scope Educational Research Foundation's evaluation of the Michigan School Readiness Program. Presenting a range of practical issues, such as the cost of program evaluations, and technically complex issues, such as validity and reliability issues associated with using teacher observation data, the chapter shares program evaluation strategies that have been used in Michigan. Michigan's two-tiered program evaluation strategy—with intensive data collected from children in a select group of programs and program quality data and child risk factor data collected from all programs—is described.

The final chapter that examines technical issues related to assessment, *Instrumentation for State Readiness Assessment: Issues in Measuring Children's Early Development and Learning* by John Love, provides a systems-level review of the elements of readiness, along with current political and educational factors that impact readiness assessment systems. Progress made in development and utilization of early childhood assessment instruments and the challenges that still remain are discussed. Finally, criteria for evaluating an individual measure, as well as a set of measures used in an assessment system, are presented.

Six significant challenges to developing technically sound and just assessment systems are discussed in the last chapter of this document. *Statewide School Readiness Assessments: Challenges and Next Steps* by Martha Zaslow and Tamara Halle presents a synthesis of the challenges outlined in previous papers and discussions at the symposium, along with examples of possible solutions and recommendations for next steps. This thoughtful summary of challenges states face provides an overview of issues that need to be addressed as wide-scale early childhood systems move forward.

The purpose of this special report is to provide a discussion of the complex issues involved in planning and implementing wide-scale assessment systems in order to guide policy and technical decisions within states currently involved in implementing wide-scale assessment systems, as well as states considering such systems. The ideas presented within the chapters represent each author's perspective, rather than the views of the volume's editors. The information presented will be helpful to early childhood state specialists in state departments of education, researchers, child advocates, and policymakers involved in designing assessment systems to collect data from large numbers of pre-kindergarten and/or kindergarten-age children. While it is increasingly clear that there is no perfect wide-scale early childhood assessment system, assessment systems that benefit both children and the users of assessment data can be developed with careful planning and consideration of the issues outlined in this report.

**Catherine Scott-Little,**
Expanded Learning Opportunities Project Director
The Regional Educational Laboratory at SERVE

**Sharon Lynn Kagan,**
Virginia and Leonard Marx Professor of Early Childhood and Family Policy
Teachers College, Columbia University

**Richard M. Clifford,**
Senior Scientist and Co-Director
National Center for Early Development and Learning
Frank Porter Graham Child Development Institute

## Reference

Saluja, G., Scott-Little, C., & Clifford, R. M. (2000). Readiness for school: A survey of state policies and definitions. *Early Childhood Research and Practice, 2*(2).

# *Assessing Young Children:*

## What Policymakers Need to Know and Do

**Sharon L. Kagan**, Teachers College, Columbia University
**Catherine Scott-Little**, The Regional Educational Laboratory at SERVE
**Richard M. Clifford**, National Center for Early Development and Learning

## Abstract

Policymakers interested in developing sound assessment systems that benefit children and programs face great challenges. Following background information about the nature of early childhood assessments, this paper provides recommended policy considerations, followed by specific policy actions that can be taken to promote technically sound and effective early childhood assessment systems.

## Introduction

Assessing children's readiness for school is an issue of mounting political, educational, and social concern. For example, in an unprecedented move, President Bush has called for the assessment of all three- and four-year-old children in the nation's Head Start program. While many states have had provisions that encourage assessments for this age population, few have moved as comprehensively or audaciously (Doherty, 2002, p. 66). Indeed, even those states that have serious intentions regarding the use of assessments for preschool-aged children have encountered significant challenges as they embark on this goal (Hoff, 2002). Are these calls for assessing young children appropriate? Under what circumstances are assessments meaningful, and how can they help children while providing policymakers the information they need? The purpose of this policy brief is to explore the nature of assessment for young children and the challenges it poses. In addition, critical considerations are offered for those policymakers wishing to advance appropriate and effective assessment policies.

### *Why Assess Young Children?*

Numerous rationales for assessing young children have been offered. First and most germane to this policy brief is policymakers' interest in knowing how many children are ready for school. Often, they assume that a blanket assessment of readiness exists, is easy to use, is culturally fair, and can be implemented with limited effort and cost. This is not the case. The second rationale is that such assessments would be able to tell us if preschool programs were doing their job (NRC/IM, 2001) based on the assumption that effective preschool programs should be judged on their ability to produce academically ready children. However, early childhood programs cater to widely diverse populations, often do not serve children on a full-time or even on a regular basis, and according to every empirical study done, do not have

the resources to do their jobs well. The third rationale for assessing young children, and perhaps the most important, is that well-constructed and well-implemented child-sensitive assessments can improve the instructional program and parents' knowledge of their children's development. The fourth and final rationale is that assessments will help programs identify children who may need additional services, as in the case of children with disabilities.

## Clarifying Assessment: Enunciating Principles

Each of these purposes of assessment differs and brings with it opportunities and challenges. Because so much confusion regarding assessment abounds—especially regarding its definitions and purposes—the National Education Goals Panel, through its expert panel on readiness assessment for young children, has articulated baseline purposes and principles to frame both analysis and action (Shepard, Kagan, & Wurtz, 1998).

- There are four primary purposes of assessment: (1) to support learning and instruction, (2) to identify children for additional services, (3) to evaluate programs and monitor trends, and (4) to provide information for high-stakes accountability.
- Each assessment should be tailored to a specific purpose and should be reliable, valid, and fair for that purpose.
- Assessments should be age- and linguistically-appropriate, both in content and method of data collection.

### *What's So Special About Assessing Young Children?*

Appropriate to student assessment in general, the principles stated above are particularly important for the assessment of young children. Because young children learn in ways and at rates that are different from that of older students, the content and procedures of their assessments must be somewhat different (Kagan, Moore, & Bredekamp, 1995). For example, young children learn best by listening, observing, questioning, and experimenting, and because they represent their knowledge by showing or talking and have limited abilities to communicate through written language, conventional paper-and-pencil tests appropriate for older students are not adequate for them. Young children's learning is also highly integrated and extremely episodic, so tests given at one point in time and focusing in one content area (e.g., numeracy or literacy) are not adequate proxies for the full scope and depth of the knowledge young children possess. It is often necessary to use multiple means of assessing children to gain an adequate understanding of their level of knowledge and skill in any given area. Young children are often inexperienced in adapting to new situations, and as a result formal testing settings do not effectively capture their development. Finally, because young children's achievements are strongly influenced by their past learning opportunities as well as their ability to learn, we cannot assume that measures of past learning are evidence of what might be learned (Graue, 1993).

## What Do We Do?

We know we need information about young children. Parents want to know how their youngsters are doing, teachers need systematic information to plan appropriate programs, and policymakers need to know the degree to which public investments in programs for young children are paying off. The question is what can and should policymakers do to honor the principles of assessment and to advance the production of necessary data? Are these legitimate reasons for conducting assessments while children are still getting ready for school reconcilable with inherent difficulties? The answer is yes. Carefully designed and executed broad-scale assessments can provide valuable information for all of these purposes. There are several important steps that can be taken to help ensure that

assessments are good for children, practitioners, and policymakers. The next section details these steps followed by policy recommendations.

***First, recognize that effective assessments for young children are not easy to conduct.*** The assessment industry that has been operative for decades with regard to older students has not, until recently, applied its talents and understandings to younger children. There are no perfect, off-the-shelf, easy-to-do assessments that will address the multiple purposes indicated above. There are assessments that may fill part of the bill, but these need to be examined for their scientific, cultural, age, and linguistic appropriateness. In addition, although *informal* assessment is the sine qua non of quality early care and education, *formal* assessment has not been part of the skill repertoire of early educators. Therefore, to implement effective formal assessment procedures, training early educators in their use and application must take place. Significant lead-time and financial resources will be necessary to develop an effective large-scale assessment system.

Given the challenges described above, it is questionable whether formal assessments can be used presently to reflect the effectiveness of investments made in preschool education. In other words, there are critical needs related to (a) the instruments themselves, (b) the training of those who use them, and (c) the uses of the data that need to be addressed. These issues and how they are handled directly predict the success of the assessment enterprise. Recognition of the complexity of the task is the first step toward development of a successful assessment system.

> ***Policy Implications:*** Federal and state governments should support efforts to plan for the development and implementation of assessments that will address the various purposes of assessments. Such planning efforts should include parents, policymakers, early educators, assessment specialists, and the public. Planning must address issues related to who will be included in the assessment; the relationship between local, state, and national accountability needs; measures to be used/developed, sampling techniques to be employed, and funding available for the assessment.

***Second, think assessment systems, not individual assessments.*** Because there is such widespread interest in assessment, it is often tempting to use one assessment instrument for several purposes. As noted above, this is not generally advisable because assessment instruments are designed with specific intentions that are not effectively transferred. The purposes of the assessments must be delineated, and the parameters of an assessment system clearly defined. Is the assessment system expected to track changes in children's condition over time? To evaluate a preschool program? To provide information to help teachers and parents work effectively with children? A combination of these purposes? Once the purposes are delineated, the structure of the system can be created and appropriate strategies implemented. It is better to accomplish one goal for assessment well than to try to do many things and end up doing all of them poorly.

> ***Policy Implications:*** Programs and/or states must think broadly, and often across traditional agency boundaries, about the nature of the assessments that are needed. Adequate resources and personnel must be made available for the conceptualization and implementation of a system over time, phasing in elements of the system.

***Third, support the development of adequate assessment instruments.*** To date, the early childhood/child development field has relied on instruments that have been developed for primary use in field trials or program evaluations. As such, many of the instruments, because of their length, technicalities, and cost, are not suitable for use for large-scale assessments across a variety of settings. New kinds of instruments, appropriate to mass use, must be created. In addition, many available instruments tend to be domain specific, with the result that not all the domains of development now associated

with readiness (e.g., physical health and motor development; social and emotional development; approaches toward learning; language, literacy, and communication; and general knowledge) are adequately addressed (Kagan et al., 1995). For example, a recent review of the literature concerning the state of our ability to predict future developmental status noted that there are better across-time predictions of children's cognitive development than of their socioemotional development (LaParo & Pianta, 2000). Instruments that address diverse cultures are lacking, as are assessments in languages familiar to and considerate of children's home language and culture. Equally important, there is some confusion surrounding the intentions of such assessment: Is the purpose of assessment to determine initial levels of English language proficiency, to determine content mastery in first and/or second language, or to assign children to appropriate instructional settings? Clear goals for any assessment effort are essential for success.

> ***Policy Implications:*** Prior to investing large sums of public dollars in the assessment of young children, federal and/or state governments should provide funding to evaluate the efficacy of current assessment instruments for the purposes needed. To the extent that appropriate assessments do not exist, public funds should be invested in instrument development.

***Fourth, design an approach to assessment that is sound and will produce reliable and meaningful results.*** Assessing young children typically requires multiple methods, including teacher and parental evaluation of children's skills and abilities. Some strategies for collecting such information are particularly sensitive to bias. When the results of assessments can have a direct impact on the children or on the teachers (e.g., they are high-stakes tests), extreme care is needed to ensure that the results are not biased by the self-interest of these informants.

> ***Policy Implications:*** Adequate resources must be committed during the design phase of any assessment system to ensure that the rights and well-being of children and staff are well-protected, while at the same time obtaining valid information to accomplish the goals of the endeavor.

***Fifth, support in-service professional development for those who conduct the assessments.*** Assessments that are used for instructional improvement, as well as assessments for tracking or program evaluation purposes, can and should be administered by teachers who work directly with children. In many cases, given very high turnover rates and limited training entry requirements, those who work with young children are not familiar with formal assessment. The lack of training also means many are not well-equipped to translate the assessment results into meaningful instructional practices. Mandated training in assessment for early educators should be considered, as a part of their preservice education where appropriate and as a part of ongoing in-service education. Training is also needed for those who are not classroom-based and are performing assessments. In all cases, assessors must understand the unique characteristics of young children and must be prepared to adapt to diverse early childhood settings.

> ***Policy Implications:*** In all legislation that mandates assessment and/or professional development, ensure that the early educators have the opportunity to learn how to assess young children and how to effectively use the data to plan programs and to report to parents. Require uniform training for any assessors who are assessing children to provide data that will be used for tracking or evaluation purposes. Professional associations and agencies governing professional training should develop guidance and formal requirements for professional development programs regarding training in assessment.

***Sixth, recognizing that the majority of early education takes place outside of formal programs, plan for the involvement of family childcare and other providers in the assessment system and in the accompanying professional development.*** Long overlooked, family childcare provides services to young children and their families, often serving as an information hub for parents. In addition, kith-and-kin care is another critical element in providing services for young children in this country. All of these providers need to be able to accurately assess young children and to use the information to inform their practice and the parents of the children they serve. Determining how best to engage this diverse group of providers, many of whom are not licensed or registered, presents a challenge that needs to be addressed if we want all of America's children ready for school.

> ***Policy Implications:*** Consider the unique situation of family childcare as well as kith-and-kin care and design assessment systems that can benefit these adults and the children they serve. To that end, a special national task force on family and relative childcare should be established to address these unique challenges.

***Seventh, make provisions for including parents and other family members in the assessment process.*** Parents and other family members spend more time than anyone else with children before they come to school; they know their children best. Parents and other family members can provide a wealth of information about children's abilities and characteristics, but they are often left out of the assessment process. Most commonly, parents are asked to fill out a cursory kindergarten registration form to provide basic information, such as where the child lives, and are not asked to provide information about the child's skills and interests.

> ***Policy Implications:*** Include opportunities for parents and other family members to provide information about children as part of the assessment process. Surveys and checklists are efficient ways for parents to provide their important perspective on a child to the kindergarten teacher for instructional assessments and for tracking or program evaluation assessment systems.

***Eighth, understand that readiness results from a combination of factors, all of which must be assessed.*** While the points above address early childhood assessment in general, it has been noted that the results from early assessments make, at best, only small to moderate contributions to the predictability of children's early school success, a conclusion that also obscures the extent to which non-child factors predict readiness (Kagan, Rosenkoetter, & Cohen, 1997; La Paro & Pianta, 2000). Non-child factors include, at a minimum, the role of the family and the nature of children's experience in early learning settings (e.g., the childcare and school contexts). To discern children's readiness, then, it is critical to examine the experiences to which they have been exposed and the nature and degree of such exposure. We need to understand the nature of the parenting children have received and the nature and quality of their preschool experiences. We also need to know the degree to which schools are ready for the unique learning needs of young children (National Education Goals Panel, 1998). These factors link to form the composite of children's readiness for school.

> ***Policy Implications:*** Policymakers should provide support for the development and implementation of readiness assessments that embrace non-child dimensions of readiness, including assessments of schools' readiness for children and communities' support for young children and their families.

***Ninth, clarify the way, and by whom, readiness information will be used and disseminated before the data are collected.*** As noted above, multiple rationales for assessing young children exist and often collide with one another. By being quite

intentional regarding the purposes of the data collection, not only can the instruments and process be designed appropriately but the collected data will also have the greatest utility. If, for example, evidence of the effectiveness of investments in preschool education is needed, it is not sufficient to garner information on assessments of child well-being only. Monitoring and tracking the status of children at kindergarten entry must be linked to data on program participation, expenditures, and quality in order to meet this goal. Moreover, if this is the goal, then clear ways of communicating information to relevant audiences must be anticipated. Discerning effective ways to report information in a timely and relevant way while not oversimplifying or distorting the data demand attention.

> ***Policy Implications:*** When calling for assessment information, be precise about the purposes of the data and the ways in which the data will be reported and used. Incorporate such information into legislation and regulation.

## Conclusion

In conclusion, it is important to remember that effective programs are grounded in effective assessment. To be effective, assessment must be done intentionally and with care if the intended results are to be achieved. Conventional assessment, including group-administered, norm-referenced standardized tests are not appropriate for young children. Similarly, it is not appropriate to use assessments developed for one purpose for others. As a result, new assessment strategies and approaches are needed. Tinkering with existing instruments or processes will not be sufficient to address the needs of young children or the needs of policymakers and administrative agencies that need and deserve the data.

When all is said and done, suitable assessments for young children are feasible and desirable, so long as the investments in their development are made. Much like the field of early education itself, assessment is commanding much attention, with the sense that capacity already exists. And much like the field itself, early childhood assessment lacks the infrastructure to support its immediate implementation. Many things can and should be done now, and not all of these will yield that data that policymakers want for the next legislative session. Rather, the assessment of young children should be regarded as an investment to be made over time. As quality care and education is requisite to young children's optimal development, so too is effective assessment requisite to quality early care and education. Both necessary and complex, neither will happen overnight or without the oversight of thoughtful, caring policymakers.

## References

Doherty, K. M. (2002). Early learning: State policies. *Quality Counts 2002, Building Blocks for Success: State Efforts in Early Childhood Education* [Special issue]. *Education Week, 21*(17).

Graue, M. E. (1993). *Ready for what? Constructing meanings of readiness for kindergarten.* Albany: State University of New York Press.

Hoff, D. (2002). Measuring results. *Education Week Special Edition, 17,* 48–52.

Kagan, S. L., Moore, E., & Bredekamp, S. (Eds.). (1995). *Reconsidering children's early development and learning: Toward shared beliefs and vocabulary.* Washington, DC: National Education Goals Panel.

Kagan, S. L., Rosenkoetter, S., & Cohen, N. (Eds.). (1997). *Considering child-based results for young children: Definitions, desirability, feasibility, and next steps.* New Haven, CT: Yale Bush Center in Child Development and Social Policy.

La Paro, K. M., & Pianta, R. C. (2000). Predicting children's competence in the early school years: A meta-analytic review. *Review of Educational Research, 70*(4), 443–484.

Meisels, S. J. (1987). Uses and abuses of developmental screening and school readiness testing. *Young Children, 42,* 4–6, 68–73.

National Education Goals Panel. (1998). *Ready schools: A report of the Goal 1 Ready School Resource Group.* Washington, DC: Author.

National Research Council and Institute of Medicine. (2001). *Getting to positive outcomes for children in child care: A summary of two workshops.* Board on Children, Youth, and Families, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Shepard, L., Kagan, S. L., & Wurtz, E. (1998). *Principles and recommendations for early childhood assessments.* Washington, DC: National Education Goals Panel.

# *A Risk Management Approach to Readiness Assessment:*

## Lessons from Florida

**Susan Muenchow**, American Institutes for Research

## Abstract

In conjunction with school readiness initiatives, many states are calling for assessments of preschool and kindergarten-age children. Drawing especially on lessons from Florida, this paper reviews the potential benefits and risks of school readiness assessment and then suggests some strategies for risk management. Principles developed by the Florida Partnership for Children to guide development of their assessment system are included. Proposed strategies to help maximize the potential benefits of readiness assessment and to minimize the risks include (a) involving child development specialists and stakeholders in the planning process, (b) developing a set of principles for the assessment system, (c) clarifying that no one instrument will fit all purposes of assessment, (d) articulating the costs of a responsible assessment system, (e) clarifying the procurement process, (f) releasing readiness data in conjunction with program quality and demographic data, and (g) making the case for state participation in a national evaluation of program effectiveness. The paper concludes that a uniform school readiness assessment is useful for providing benchmark and trend data on the status of young children across counties or school districts but cannot substitute for program evaluation.

## Introduction

For many years, early childhood professionals discouraged the practice of formal readiness assessment for young children. Six professional organizations, including the National Association for the Education of Young Children and the Association for Childhood Education International, issued a joint statement in 1986 discouraging the use of standardized testing for preschool children (Saluja, Scott-Little, & Clifford, 2000; International Reading Association, 1986). Within the last few years, however, a number of states have begun to call for assessments of preschool and kindergarten-age children in conjunction with school readiness initiatives. Without abandoning their concerns about the risks associated with readiness assessment, early childhood professionals are struggling with how to respond to legitimate requests for information. The question has become not whether to assess young children, but how to do so in a manner that is developmentally appropriate and that causes no harm. Beginning with an explanation of the use of the term "assessment," this paper will review the potential benefits and risks of school readiness assessment and then suggest some strategies for risk management. The paper focuses on the experience with readiness assessment in Florida where the author served as the first director of the Partnership for School Readiness.

*The question has become not whether to assess young children, but how to.*

## Use of the Term "Readiness Assessment"

For purposes of this paper, the use of the term "readiness assessment" will not be limited to a single snapshot of the child but may include various components, of varying intensity, and conducted by a range of personnel. It is important to explain the broad use of the term at the outset because some of the misunderstandings regarding assessment stem from confusion over the meaning of the term. Components of an assessment system may include a developmental screening to identify children in need of further evaluation for possible developmental delay, observation of children over a period of time for purposes of instructional improvement, and the use of one or more assessment instruments for purposes of program evaluation. In discussions of accountability, policymakers often use terms such as "assessment" and "screening" interchangeably; their primary interest, however, is in child assessment as a means to program evaluation. In this paper, "readiness assessment" refers to a system that may include all of the above components—developmental screening and evaluation, instructional assessment, and program evaluation—so long as the purpose of each system component is clear and its use appropriate to its purpose.

## Potential Benefits and Purposes of Readiness Assessment

*Purposes of readiness assessment*
- *Identifying children with special needs*
- *Improving instruction*
- *Evaluating programs*
- *Obtaining benchmark data*

As has been well-articulated by others (Shepard, Kagan, & Wurtz, 2001), there are at least four potential benefits or purposes of readiness assessment. The purposes will be reviewed here as follows: (1) identifying children with special needs and health conditions, (2) individualizing and improving instruction, (3) evaluating program effectiveness, and (4) obtaining benchmark data on the status of children at the local, state, and community level.

### *1. Identification of Children with Special Needs or Health Conditions*

The first and clearest reason for readiness assessment is to identify children with special needs or health conditions. Child development specialists have long expressed the importance of early identification and treatment of special needs. In the wake of the popularization of research on the development of the brain, state legislators have become interested in the concept of certain "windows of opportunity" for the development of vision, hearing, emotions, and language, and the idea that the first five years of life have a lasting impact on a child's physical, emotional, and intellectual development. State laws requiring universal hearing screening for newborns and vision screenings or exams for children prior to kindergarten entry represent one positive response to the interest in early identification of special needs and health problems. Some states are also enacting requirements for screening all children enrolled in state-funded early care and education programs in order to determine if the children should be referred for evaluation for possible developmental delay. Although identification and treatment of special needs should occur long before kindergarten entry, the design of a readiness assessment system provides the opportunity to include screening components at birth, during well-child visits, and during participation in a variety of early education and care experiences. Furthermore, a screening upon kindergarten entry offers an opportunity to identify conditions that, for whatever reason, have not been previously detected.

### *2. Individualizing and Improving Instruction*

The second reason for embracing readiness assessment is to seize the opportunity to individualize and improve instruction at both the preschool and kindergarten level. Parents of children enrolled in preschool programs want feedback on how their children are developing and what can be done at home and in the program to help a child's

physical, emotional, cognitive, and language development. Teachers in early childhood programs welcome assistance in how to better understand the children in their care. When a state law requiring developmental screening of all children enrolled in state-subsidized childcare was enacted in Florida, childcare administrators and teachers at first complained that there were no new funds to implement the program and that the workforce was not up to the task. But after piloting the screening (Ages & Stages) in eight areas of the state, the same administrators and teachers wanted to continue the screening even if new funds were never provided. The reason was that they saw the effort to conduct the screening not only as a way to identify possible disabilities but also as the best training in the observation of young children that many of the teachers had ever received. Similarly, teachers piloting the Desired Results Developmental Profile in California reported the observation tool helped them to think in terms of the "whole" child, across developmental domains. In short, as Meisels and Atkins-Burnett (2000) point out, the assessment (in this case, a simple screening) cannot be separated from the intervention or early education and care. The value of the assessment lies not only in the information obtained on the child's status but also in the fact that the teacher begins to "know" the child well enough to improve the early care and education for the child.

Similarly, as children enter and progress in kindergarten, they benefit from the teacher spending some time assessing their individual learning styles, strengths, and weaknesses. The average expenditure per child for a public K–12 education in the United States is more than $92,000 according to the National Center for Education Statistics (2001), based on 13 years of education at the average annual expenditure of $7,079 per student. Imagine embarking on any other comparable investment of resources without first taking the time to conduct some type of initial assessment of the person one is attempting to assist. Just as a surgeon would not start an operation without first diagnosing and assessing the patient's overall health condition, a teacher should not start a year of instruction before having an opportunity to get an initial reading of each child's social-emotional, cognitive, motor, and language development. For parents and teachers, instructional improvement may be the most important reason to support readiness assessment.

### *3. Holding Early Childhood Programs Accountable*

The primary current interest of national and state policymakers in school readiness assessment, however, is in accountability—evaluating the effectiveness of early childhood programs in preparing children for school and holding the programs in some way accountable. The interest in readiness assessment can be viewed as just the first step in the broader effort to hold publicly funded education for all age groups accountable. More specifically, the call for readiness assessments arises from the ongoing debate in the United States regarding whether it should be a national priority to invest in early childhood programs and, if so, for which children and at what level of expense. Although investments in early childhood programs have increased substantially at both the national and state levels in recent years, astute policymakers, including advocates as well as opponents of increased investment, realize that serving all currently eligible children, much less all children, would cost a great deal more. Legislators, told that investments in early intervention for disadvantaged children will reduce school failure, want to know if the programs are effective. While researchers and advocates may wonder why the outcomes of national evaluations involving assessments of children would not be sufficient to address these questions, all politics is local and legislators respond best to the evidence found closest to home. Researchers are therefore in the position of being asked to address the policymakers' "need to know."

### *4. Holding National, State, and Local Policies Accountable*

Another reason to support readiness assessments is to obtain benchmark data on the status of young children, state-by-state, county-by-county. The data can help determine whether the status of school readiness is improving and how it compares to the status of children in

other states and localities. Annual profiles, such as the Annie E. Casey Foundation's *Kids Count*, track important child-health outcome indicators such as low birth weight, infant mortality, and immunizations. *Kids Count* also tracks fourth-grade reading and math skills. However, for benchmarks between infancy and the mid-point of elementary school, *Kids Count* and other progress reports are generally forced to rely on process indicators (e.g., percentage of children enrolled in early education and care programs, accreditation status, etc.). The addition of school readiness outcome measures, assuming they can be agreed upon and administered reliably, would provide a missing link between maternal and child-health outcome indicators and mid-elementary school-age indicators. In a recent article, Shepard et al. (2001) state that it is possible to develop such measures beginning at age five as part of a comprehensive early childhood system to monitor trends. The combination of demographic information, outcome indicators, and process measures across the age span of children would be a powerful tool to hold policies—and policymakers—accountable.

## Risks of Readiness Assessment

While there are many potential benefits associated with readiness assessment, there are just as many risks. Reasons for caution range from the unintended consequence of drawing inappropriate conclusions from misleading data, to diverting already scarce resources from program expansion and improvement to assessment, to denying children access to kindergarten, to punishing early childhood programs for serving the most disadvantaged children.

*Risks of readiness assessment*
- *Inappropriate conclusions*
- *Diversion of scarce resources*
- *Denying children access to kindergarten*
- *Punishing programs for serving the most disadvantaged children*

### *1. Drawing inappropriate conclusions from misleading data*

The clearest risk associated with readiness assessment is simply that the data released for accountability purposes will be improperly gathered and used to draw inappropriate conclusions. Examples of misuse of school readiness data already exist and perhaps offer the best illustration of the need for assessment and evaluation experts to get involved in readiness assessment (i.e., to improve it).

According to a recent survey by Saluja et al. (2000), 13 states currently collect readiness data, but only a few require school districts to use the same, much less a standardized, instrument. There is no guarantee, however, that states allowing localities to use different instruments will not analyze, report, and use the data as if it were based on the same instrument. In 1996, for example, Florida, as part of a broader school accountability initiative, began collecting data on school readiness. Each school district was given a 16-item checklist, and kindergarten teachers were asked to assess readiness during the first three weeks of school. To be considered "ready," a child must score 75% or better on the checklist items. Each school district was allowed to select its own method for measuring the items on the list. The 67 districts therefore chose a variety of instruments—Brigance, the Child Observation Record, the DIAL-R, locally developed instruments, or the checklist itself, which was not designed to be a screening instrument. Most districts were not able to train the teachers in how to administer the instruments selected. Despite the obvious flaws in the process, school districts had to dutifully report the results each year to the Department of Education. A summary of the readiness scores by district was then released to the legislature, to program evaluation staff in the Governor's office, and to anyone else upon request. Information indicating the variety of instruments used to measure readiness was rarely, if ever, included in the summaries regarding the readiness results.

Further compounding the above problems in the readiness data, the state began using the scores to assess the effectiveness of subsidized childcare. Several years ago, as a part of a new performance-based approach to budgeting, the state set a requirement that at least 80% of children enrolled in state- and federally-funded care must be "ready" for school.

The ultimate purpose of setting the standard for the outcome measure would be to use the information for performance-based budgeting. To satisfy the request for information, readiness scores for children entering kindergarten were matched with the database for children previously enrolled in subsidized childcare. Children's readiness status was then further analyzed by the type of subsidized care in which they had been enrolled—informal care, voucher or contracted family childcare, or voucher or contracted center care.

Fortunately, this story has a happy ending. To this author's knowledge, the above data from locally selected instruments were never used for performance-based budgeting purposes or for any other high-stakes purpose. With the passage of Florida's School Readiness Act in 1999, the state mandated the development of a uniform assessment for all children entering kindergarten. (Although the statute uses the term "uniform screening," the purpose is not only to identify children with possible developmental delay but also to assess the broader "school readiness" of entering kindergartners in the areas of health, cognitive, social and emotional, and language development.) Funding has been made available to purchase the assessment materials statewide and to train teachers in the use of the materials. The Florida Department of Education has clearly articulated that there are problems with the old system of measuring readiness, including the reliability and validity of the locally selected instruments, with inconsistencies in administration, and with inappropriate use of reported data.

Let us suppose, however, that the above school readiness data *had* been used for performance-based budgeting purposes. The following are just a few of the conclusions a reader might draw from the summaries since 1996 of school district data based on locally selected instruments:

- The readiness status of children in Florida ranges dramatically, from as low as 41.3% "ready" in one county to as high as 95.8% in another (1996–1997 survey).
- In some of the poorest counties in the state, with the lowest levels of maternal and child health and the highest dropout rates, nearly all children nevertheless enter school ready to learn (all surveys since 1996–1997).
- Informal subsidized care is the superior form of care, based on the data that 100% of the children who received state-subsidized informal care were ready for school in 22 counties (no matter that in many of these counties only one or two children were actually receiving subsidized informal care, 1999–2000 school readiness data).

In summary, the data from the locally selected instruments were not comparable across districts, were frequently not based on reliable and valid instruments, and in many cases were not collected from a large enough group of children from which to draw any conclusions. To use the above data for purposes of performance-based budgeting would be a disservice to the programs, the children enrolled, and the taxpayers.

### 2. *Diverting resources from program expansion and improvement to assessment*

Even if readiness assessments are well designed and used appropriately, there is some concern that the cost of the assessment(s) will absorb already scarce resources for expansion and improvement of early childhood services. The per-child cost of the assessment battery, including administration, in national studies such as FACES or ECLS-K (Early Childhood Longitudinal Study—Kindergarten Class of 1998–1999) is more than $400 per year. While this is a manageable expenditure for a sub-sample, it would probably be excessive if applied to every child enrolled in publicly funded school readiness programs (i.e., more than 15% of the expenditure for the program in many states). Even when the assessment

is limited to a simple developmental screening component, there is also the issue of time burden placed on the assessors, particularly if they are teachers. Given teacher turnover, there is the need for ongoing training. As stated above, teachers benefit from involvement in the screening of the children in their charge because they gather knowledge that enables them to improve the instruction for the child. But clearly there is a limit to how much time teachers will be able—and willing—to divert to assessment from other important classroom responsibilities.

There is also the very real possibility that state legislatures will use readiness assessments as an excuse not to finance program expansion or improvement until the results of the assessment are available. The Florida School Readiness statute, for example, provides for coalition incentive grants for county coalitions that can prove they are improving the school readiness of the children in their service area. In the 2000–2001 legislative session, leaders made clear that they were awaiting the outcome of the school readiness data before providing significant new funds for program improvement. Given the time required to implement a responsible assessment system, children may have to wait several years until the outcome data are in to support program improvements. There may also be a tendency to inflate the scores if they are tied to financial rewards for programs.

### *3. Denying children placement in kindergarten*

Another concern is that readiness assessments will be used for denying children entrance to kindergarten, or that they will be used for requiring children to complete an extra year of school between kindergarten and first grade. While 26 states once conducted readiness screening or testing, with several states using the results to delay school entry, most states have now discontinued the practice (Shepard, Taylor, & Kagan, 1996; Saluja et al., 2000). There is recognition that young children are difficult to assess and that the assessment results are therefore not sufficiently reliable to justify denying any child the benefit of a year of publicly financed education. Nevertheless, pendulums can swing both ways. Despite initial recommendations by the Florida Partnership for School Readiness that kindergarten assessment scores only be reported in aggregate to the state, it was later determined that the scores would be part of a student's record. Furthermore, once the teacher knows a child's performance on an assessment, he or she will be armed with the information to recommend retention in kindergarten.

### *4. Punishing programs for serving the most disadvantaged children*

Finally, there is great concern that using readiness assessments for accountability purposes will inadvertently punish the programs that serve the most disadvantaged children, or, put another way, discourage publicly funded early childhood programs from serving children who most benefit from being in the programs. For example, if a program's level of state funding were determined by readiness outcomes, the program might accept fewer children with special needs. This is a very real possibility if preschool and other early childhood programs are evaluated solely on the basis of exit scores from preschool programs and/or entry scores in kindergarten, as opposed to the learning gains children make from entry into preschool programs until entry in kindergarten.

Two states currently involved in the development of assessment systems seem to be well aware of the danger. Ohio, which has been tracking Indicators for Success since 1997 for 25,000 children using Galileo software, is including entrance as well as exit data from Head Start and other early childhood programs. Similarly, Florida's legislation calls for both pre- and post-assessment of children in publicly funded preschools. Florida's legislation also calls for a longitudinal evaluation to track the performance of children through third grade who have been enrolled in early childhood programs and those who have not. Nevertheless, the proof will be in the implementation. The 2001 Florida Legislature provided the funds

necessary to implement the uniform kindergarten assessment (again, in the Florida law, this is called a "uniform screening") but not those needed for the longitudinal study. Policymakers find it difficult to understand why the uniform kindergarten assessment, with scores tracked back to the early childhood program in which the child was enrolled, cannot substitute for the longitudinal evaluation. Once the kindergarten assessment scores are released to the public, they will take on a life of their own.

## A Risk Management Approach to Readiness Assessment

Given the potential benefits of school readiness assessment, how can states minimize the associated risks? Several strategies will be proposed to help maximize the potential benefits of readiness assessment and to minimize the risks. These include: (1) involving child development specialists and stakeholders in the planning process, (2) developing a set of principles for the assessment system, (3) clarifying that no one instrument will fit all the purposes of assessment, (4) articulating the costs of a responsible assessment system, (5) clarifying the procurement process, (6) releasing readiness data in conjunction with program quality and demographic data, and (7) making the case for state participation in a national evaluation of program effectiveness.

Strategies to maximize benefits and minimize risks:

- *Inclusive planning process*
- *Guiding principles*
- *Multiple instruments when there are multiple purposes for assessment*
- *Costs articulated*
- *Procurement process clarified*
- *Data released strategically*
- *Advocation for multi-state evaluations*

### *1. Involving child development specialists and stakeholders in the planning process*

Perhaps the single most important step to maximize the benefits and limit the risks of school readiness assessment is to involve child development specialists and stakeholders in the planning of the system. Their guidance is not only needed in the design of the system but also to provide feedback at various points during the development and implementation of the system.

Enacted in May 1999, the Florida School Readiness Act required the newly created Florida Partnership for School Readiness "to prepare and submit to the State Board of Education a system for measuring school readiness" by July 2000, only six months after the Partnership staff of three full-time staff was actually assembled. The timeframe was formidable, and the stakes were high. First, the legislation called for a "uniform screening" of all children entering kindergarten, which would have the merit of providing a state mandate to identify any children with previously undetected developmental delays or health problems. The legislation also anticipated that the same screening tool would serve as a broader assessment to measure children's readiness status in language, social-emotional, and cognitive development. Although the legislation required a longitudinal evaluation for purposes of determining program effectiveness, it also required the uniform assessment of *children* entering the kindergarten track back to the early care and education programs in which the children had been enrolled, suggesting that the sum of the individual child assessments and the program evaluation were to be one and the same.

Faced with this challenge, the Partnership was fortunate to obtain the help of a lead consultant, Dr. Sharon Lynn Kagan, who serves as an advisor to the National Education Goals Panel. In addition, the Partnership assembled a state-level workgroup composed of child development specialists, early childhood leaders, assessment experts, kindergarten teachers, disabilities specialists, health professionals, and state agency officials. The workgroup then considered assessment strategies in a number of other states and had presentations from other consultants on such topics as the Desired Results Project from California, Work Sampling, and a comparison of the Head Start Performance Measures with school readiness assessment.

Without the help of this workgroup, the Partnership would not have had the time to develop a responsible proposal for the assessment system within the statutory timeframe. Furthermore, the status of the workgroup members, not to mention that of the nationally recognized lead consultant and the various presenters, added credibility to the recommendations. The workgroup, composed primarily of persons outside government, provided a defense against those who thought the assessment should indeed be used for purposes of delaying some children from entering kindergarten and establishing a special program for them. Finally, the workgroup developed a knowledge base to provide ongoing support for the proposed system.

In retrospect, although the Florida workgroup on readiness assessment included kindergarten teachers and education agency representation, it would have been further enhanced by the inclusion of representatives of the school administrators' association. Because schools would ultimately be charged with conducting Florida's uniform assessment of children entering kindergarten, it would have been helpful to have their suggestions regarding the logistics of the screening implementation from the outset.

### *2. Developing a set of principles for the assessment system*

The second most important step in a risk management approach to readiness assessment is to develop a set of guiding principles. With the help of the consultants, the workgroup proposed—and the Partnership Board later adopted—a list of principles that are summarized here as follows:

**Principles Regarding Readiness**

- School readiness is the match between the condition of young children as they enter school and the capacity of schools to educate all children.
- Standards, pedagogy, programs, and assessment instruments must be based on six domains of children's development: (1) physical health, (2) approaches toward learning, (3) communication and language development, (4) social/emotional development, (5) motor development, and (6) cognitive development and general knowledge.

**Principles Regarding the Use of Data**

- Data collected from assessments should bring about benefits for the children from whom data is being collected.
- Data collected from a uniform screening or assessment should not be used for high-stakes purposes such as the retention of individual children or the addition of a year before first grade. The only criterion to be used for entry to kindergarten is chronological age.
- Data on children's status should be considered preliminary until the assessment system has been piloted and well implemented.

**Principles Regarding the Assessment System**

- To accomplish the multiple intents specified in Florida's school readiness legislation, a system of assessment should be adopted for purposes of identifying children with potential developmental delays, instructional improvement, and program evaluation. It is understood that no one instrument can be used to meet all these needs.

**Principles Regarding Assessment Instruments and Process**

- All assessment instruments must be able to accommodate the linguistic needs of children in major language groups (meaning that the instruments will be

available in English and Spanish and, for other languages, the school will attempt to identify an interpreter to assist with the screening).

- All assessments should incorporate data from different sources over time.
- The uniform screening or assessment should be brief, easy to administer, and affordable. Results should be reported in ways that parents and citizens can easily understand.

The principles developed more than a year ago have stood the test of time. They have provided a strong foundation from which the Partnership could make recommendations for the assessment system. For more detail on the principles, see the Final Report of the Workgroup on School Readiness Assessment, *School Readiness in Florida: Strategies for Defining, Measuring, and Advancing Children's School Success* (Workgroup on School Readiness Assessment, 2000).

### *3. Clarifying that no one instrument will address all the purposes of readiness assessment*

From the standpoint of the experience in Florida, the single most important provision in the set of principles was the recognition that readiness assessment involves different purposes and the same instrument designed to meet one purpose (e.g., the uniform kindergarten assessment) is not adequate to meet another, such as program evaluation. As suggested at the outset of this paper, many policymakers use the terms "screening," "assessment," and "evaluation" interchangeably. By establishing the principle that there is no "one-size-fits-all" approach to assessment, the Florida workgroup set the stage for a comprehensive approach. The uniform assessment, including health and developmental screening, should be relatively brief, affordable, easy to administer, and already field-tested. For purposes of instructional improvement, there should be an ongoing observational assessment. For purposes of accountability, the results of the kindergarten assessment provide benchmark or trend data on the status of children across counties or school districts. This type of data can provide general guidance to policymakers on the status of young children without evaluating the effectiveness of any one program. Finally, for purposes of program evaluation, only a more comprehensive battery of measures, including information on program quality based on standardized environmental ratings, information on the child's family, and information on the child's status upon entry into a preschool program as well as upon exit and again in third grade, would be sufficient. It would not be feasible to use this battery on all children enrolled in publicly funded programs but only on a statistically valid sample.

### *4. Clarifying the costs of an assessment system*

As new states take on the task of readiness assessment, they could benefit greatly from advice on the costs of relative approaches. For example, one reason why policymakers may be tempted to try to use one instrument, such as a one-time screening, to satisfy multiple purposes of a readiness assessment system is that they may think this is the simplest, least expensive route to follow. However, establishing a system for the purpose of holding early childhood programs accountable requires assessing every child and may be the most expensive approach. First, there is the cost of the instrument itself, and because it is difficult to find any single instrument that covers all the appropriate developmental domains, multiple investments may be required. Second, in order to attempt to ensure the reliability of the results, there is the ongoing investment required to train the assessors (most frequently teachers) in the use of the assessment instrument(s). Even the simplest assessment instruments frequently require one to three days of training (Niemeyer & Scott-Little, 2001). Observational instruments suitable for the purpose of improving instruction usually require more training and ongoing support for the teachers involved. Third, if the purpose is to evaluate program effectiveness, there is the cost (and difficulty)

of establishing a data collection and management system that will reliably track the enrollment of all children, or even of those enrolled in publicly funded early care and education settings, through kindergarten. Because so much publicly funded childcare takes place in family childcare or exempt settings, this is a formidable task.

In summary, if the primary purpose is to assess program effectiveness, the cost of a universal kindergarten assessment, administered responsibly, may approach the cost of a longitudinal evaluation applied to all children. Before embarking on such an investment, policymakers need information on alternatives that may be less expensive, such as the use of uniform assessments for the purpose of obtaining benchmark or trend data, and a longitudinal evaluation of a much smaller sub-sample of children for the purpose of determining program effectiveness. With better information up front on the costs and considerations involved with the various types of assessment, states will be in a better position to address how best to hold programs accountable.

### *5. Clarifying the procurement process*

Before enacting legislation requiring readiness assessment, it is important to identify the funds available to pay for the assessment and to clarify which agency will be in charge of actually purchasing assessment instruments. Florida's system would have been up and running in the summer of 2002 had there been clarity regarding which agency (i.e., the Partnership or the state's Department of Education) would be in charge of conducting Request for Proposals to select the instruments and to proceed with the purchase.

### *6. Releasing readiness assessment data*

Once school readiness assessment data are available, it would be helpful if they could be released in conjunction with other data regarding program quality, family income, school dropout rates, etc. Releasing the readiness data as part of a package of information would help guard against simplistic interpretations and contribute to more thoughtful analysis.

It is also important for researchers and state agency officials to communicate problems in data to program evaluators in the legislature and the Governor's office. When the Florida Partnership for School Readiness explained to evaluation staff in the legislature and the Governor's office the problems with the old system of locally selected instruments for school readiness measurement, they stopped using the data for program outcome measures, even though the new readiness assessment system was not yet in place. It is in no one's interest to make important policy decisions based on obviously flawed data.

### *7. Making the case for state participation in a national evaluation of program effectiveness*

For purposes of program accountability, it would be extremely helpful if each state were not in a position of reinventing the wheel. As stated above, all politics is local, and policymakers have a legitimate interest in obtaining local information regarding the effectiveness of state-funded programs. At the same time, the expertise and resources necessary to design and implement a comprehensive, longitudinal evaluation are beyond the capacity of many states.

A suggested strategy would be for interested states to come together to participate in a joint evaluation. The best-case scenario would be a rigorous, independently implemented research study using standardized tools and specially trained assessors. Funded by a challenge grant from a federal agency and/or private foundations, the participating states might contribute at least a portion of the funds they would have spent on a state evaluation to participate in the multi-state study. With the advice of the participating states, the study team, coordinated by a neutral party, would determine the design, the instruments to be used, and the techniques for obtaining a comparable sample from each

state. They would also train the assessors. In this way, states would receive the state-specific information they need for accountability purposes, and they would also obtain data that were comparable across states. This approach would also maximize the benefits of the always limited funds states are able to invest in evaluation.

## Conclusion

In conclusion, there are undeniably as many risks as benefits associated with school readiness assessment. However, with proper planning, it is possible to avoid many of the possible pitfalls. Given that some of the most feared consequences have occurred without the benefit of assistance of child development and assessment specialists, it is prudent for them to help guide the readiness assessment process.

## References

International Reading Association. (1986). Literacy, development and pre-first grade: A joint statement of concerns about present practices in pre-first grade reading instruction and recommendations for improvement. *Childhood Education, 62*(2), 110–111.

Meisels, S. J., & Atkins-Burnett, S. (2000). The elements of early childhood assessment. In J. P. Shonkoff & S. J. Meisels (Eds.), *Handbook of early childhood intervention.* Cambridge: Cambridge University Press.

National Center for Education Statistics, U.S. Department of Education. (2001). Table on Estimated Student Membership, Number of Teachers, Revenues, Expenditures, and Pupil/Teacher Ratios, for Public Elementary and Secondary Schools, by State, for Grades Prekindergarten Through 12: School Year 2000–01/Fiscal Year 2001, Washington, DC: U.S. Department of Education. Retrieved from **http://nces.ed.gov** under Quick Tables and Figures.

Niemeyer, J., & Scott-Little, C. (2001). *Assessing kindergarten children: A compendium of assessment instruments.* Greensboro, NC: SERVE.

Saluja, G., Scott-Little, C., & Clifford, R. M. (2000). Readiness for school: A survey of state policies and definitions. *Early Childhood Research and Practice, 2*(2).

Shepard, L., Kagan, S., & Wurtz, E. (2001). Principles and recommendations for early childhood assessments. *The State Education Standard, 2*(2), 5–12.

Shepard, L., Taylor, G. A., & Kagan, S. L. (1996). *Trends in early childhood assessment policies and practices.* Report prepared for the National Education Goals Panel. Washington, DC: National Education Goals Panel.

Workgroup on School Readiness Assessment, Final Report. (2000). *School readiness in Florida: Strategies for defining, measuring, and advancing children's school success.* Adopted by the Florida Partnership for School Readiness, June 27, 2000.

# *Assessing School Readiness*

## System Design Framework and Issues

**Gary T. Henry**, Andrew Young School of Policy Studies, Georgia State University

## Abstract

Starting school ready to learn has been a prominent goal for education since it was adopted by state and federal leaders in 1989. However, only recently have states begun to develop and implement ways of directly assessing school readiness. In this paper, a number of purposes for conducting readiness assessments are reviewed to assist readers in selecting their highest priorities. A clear, guiding purpose is essential in making good decisions about the assessment system. One such purpose is discussed in greater detail—informing the public and policymakers about the overall adequacy of society's cumulative investments that prepare children for learning and development through their school experiences. Ten issues that should be considered in the development of the assessment system are raised and discussed to assist those who are guiding the development of readiness assessment systems.

## Introduction

Every day parents of young children have the opportunity to observe their children and inform themselves about their children's readiness for school. In the course of daily events, teachers, administrators, and the complement of other professionals who interact with young children observe the children and obtain information that can inform their judgments about the children's readiness for school. Those with direct access to young children can use their natural sense-making abilities to assess the readiness of children for school. However, those without direct access are limited to assessing the children with whom they come in contact. Moreover, their observations may focus on characteristics of the children that are not directly related to school readiness and the informal judgments they form can be inaccurate. Others who need accurate information about school readiness and how it is changing, but interact with children on an infrequent basis, may have more deeply fallible methods for reaching conclusions about school readiness.

Assessment systems are types of evaluation approaches which rely on systematic inquiry techniques to correct biases of our natural sense-making efforts and extend the reach of the observations that any individual can make (Mark, Henry, & Julnes, 2000). Like eyeglasses and microscopes, assessment systems are tools for assisted sense-making. Assessment systems are systematic inquiry methods that enable us to improve the accuracy of our descriptions

30 BEST COPY AVAILABLE

of the world around us: in this case, the accuracy of our descriptions of the readiness of children for school.

*Like eyeglasses and microscopes, assessment systems are tools for assisted sense-making.*

Assessment systems for young children can be put into place for a variety of purposes. Traditionally, assessments of young children have emphasized purposes such as screening for disabilities, understanding the developmental progress of individual children, or providing feedback to teachers concerning the collective development of their students (Katz, 1997). While these purposes are important and need to be met, new purposes have come to the fore with the adoption of school readiness as a national goal (Shepard, Kagan, & Wurtz, 1998). Authors have begun to include information for monitoring trends, accountability, and program evaluation in the lists of purposes for readiness assessments (Shepard et al., 1998; Horm-Wingerd, Winter, & Plofchan, 2000). In defining purposes beyond the classroom or collection of individuals who are concerned about a particular child, these authors have expanded the list of legitimate and important purposes and opened up the dialogue for expanding the collection of systematic information at the state level.

For the sake of clarity, I would like to focus on the evaluative and accountability purposes for readiness assessments. Mark et al. (2000) outline four purposes that can be served by evaluations such as state assessment systems:

1. Assessment of merit and worth
2. Organizational or program improvement
3. Oversight and compliance
4. Knowledge development

An evaluation that assesses merit and worth is an attempt to measure the overall value of a program, a policy, or a series of investments for the society. A primary reason to implement state readiness assessments is assessing the merit and worth of society's overall investments, including policies, programs, and social supports, that prepare children for learning and development through their school experiences. Readiness assessments could inform overall judgments or conclusions about the cumulative investment in young children made by a state or the nation as a whole. In contrast to a *program* evaluation serving the same purpose, readiness assessment systems cannot easily inform judgments about the effects of a specific program. These systems change the level of focus from a single program to that of the constellation of programs, policies, and private investments that are made on behalf of children younger than six. Designing these systems requires us to make a cognitive leap from the level at which most evaluation efforts are focused (program effects) to a more abstract level (aggregated effects of public and private investments). The data from readiness assessment systems inform us about the benefits and consequences that result from the cumulative investments that are being made on behalf of young children, in a manner that is consistent with the theory of environmental or ecological influences on child development. Moreover, readiness assessments, which have much in common on this score with other state educational assessments and standardized tests, seek to describe rather than attribute cause (Mark et al., 2000). Although assessments can be used to estimate the effects of systemic reforms (Bloom, 1999; Henry & Rubenstein, 2002), it is very difficult to reach confident causal conclusions about the effectiveness of individual policies or programs from these assessment systems (Harkreader & Henry, 2000).

*Readiness assessments could inform overall judgments or conclusions about the cumulative investment in young children made by a state or the nation as a whole.*

A primary evaluative purpose for assessing readiness is assessing the merit and worth of or informing conclusions about the overall investment in young children, and it is much less likely to serve any of the other three purposes. The public support for children under six is largely fragmented with no institution, organization, or program having the overall responsibility to ensure that no child falls behind. Therefore, readiness assessments cannot

pinpoint organizational or programmatic deficiencies for improvement purposes. At a strategic or policy level, readiness assessments can provide information on the results of gaps in the current pattern of investments. However, readiness assessments offer little organizational or program improvement guidance, because the information can point to problems or needs but not to specific means of addressing those needs. The fragmented and diffuse nature of the assignment of responsibilities for child development also limits the utility of readiness assessments to meet traditional accountability requirements including oversight and compliance. There are no readiness compliance standards from which to judge non-compliance in the implementation of policies, rules, or regulations that are most often the focus of oversight evaluations. Readiness assessments can provide data to enhance knowledge by developing classification systems or testing theories, but these purposes are generally secondary in the view of evaluation sponsors. In spite of these limits on the use of the data, it is possible to justify the expenditures required for these systems on the basis of the need for information to assess the merit and worth of the investments made in young children.

Judgments about the adequacy and types of investments in the development of young children can influence attitudes and actions in the following very important ways:

- Increase the salience of children's issues among the public
- Provide the occasion for a focusing event (Kingdon, 1995), like the release of SAT scores or state educational assessments, that will raise the issue of the readiness of young children for school to the public agenda
- Provide a rationale and focus for institutional actions and elite deliberations (committee hearings, public hearings, Blue Ribbon panels, etc.) that can increase the attention to and legitimize the place of children's issues on the public agenda
- Justify continued and increased expenditures to meet the developmental needs of young children
- Assess gaps and needs for specific types of investment in young children
- Set societal expectations for the performance of young children that would make them fully ready to benefit from schooling
- Establish a baseline and periodic evidence of the trends in the readiness of preschool-age children
- Support efforts to build the civic will needed to adopt innovative policies for young children
- Justify adoption of new policies when the conditions of young children do not meet societal expectations

Indeed, it is quite plausible that the lack of systematic, credible evidence about children's readiness for school has inhibited the ability of children's advocates and policy experts to influence widespread adoption of innovative children's policies, especially at the state level. For example, after almost ten years of operation, Georgia stands alone in providing state-supported, full-day, developmentally oriented instruction and supervision for four-year-olds whose parents choose to enroll them (Cauthen, Knitzer, & Ripple, 2000).

We can contrast this with K–12 education where assessments are commonplace and policy innovation diffusion has occurred from one state to another (Mintrom, 1997). The current round of education reform, which commenced in 1983 with the release of *A Nation at Risk*, has not abated for nearly 30 years. During this time, public and policymaker interest in education reform is fueled every few months by the release of state assessments, the National Assessment of Educational Progress, and SAT scores

(even though the statewide averages provided by the College Board are potentially misleading, because the students who choose to take the test in each state represent a biased sample of high school students). The K–12 assessments bring the policy problems of the education system into clear focus in the minds of the public and policymakers. Evaluation research provides justification for the solutions, such as targeting early elementary grades for lower class sizes in the STAR and WISE evaluations (Mosteller, 1995; Molnar, Smith, Zahorik, Plamer, Halbach, & Ehrle, 1999). While the problems at which these initiatives are directed are plausibly, but not unquestionably, better addressed by meeting the needs of younger children where solutions have also been shown to work (Reynolds, 2000; Reynolds, Temple, Robertson, & Mann, 2001; Peisner-Feinburg & Burchinal, 1997; Barnett, 1995; Schweinhart & Weikart, 1998), there are no comprehensive indicators of the discrepancy between performance and expectation prior to the third or fourth grade. State-level information about the readiness of children for school that provides a comprehensive (all children and aspects of readiness) and valid description of the well-being of young children is largely absent, with birth indicators providing most available indicators of child well-being for children under the age of six. This implies that there are no indicators to contrast with societal expectations for children's readiness or to allow us to decide whether our investments in young children are paying off in the way that we as a society would expect until children reach the third or fourth grade. Therefore, we have no systematic evidence to provoke society to do more or to show us if doing more would increase readiness and reduce the problems associated with the lack of readiness.

*The K–12 assessments bring the policy problems of the education system into clear focus in the minds of the public and policymakers.*

School readiness assessments can provide a valuable service for society even if their only purpose is to assess the merit and worth of the constellation of investments being made on young children. Of course, it is possible to meet multiple needs, and some state assessment systems will be designed to cover a variety of purposes. However, to have the highest probability for success, the designers of state assessment systems should discipline themselves to the one or two purposes that have the highest priority (Mark et al., 2000), one of which should be to inform overall judgments about the pattern of public, private, and social investments in young children. It is possible that other purposes will be accomplished in the process, and during the design phase, it is possible to consider tweaking the systems to take another purpose into account. However, developing a clear sense of priorities offers the best chance for the system to realize its purpose.

Having a clearly defined purpose will provide those designing and implementing readiness assessment systems with the first-order criterion for making decisions about the issues that must be confronted in the process. The following ten issues presented below will be addressed in the design phase of the development of the assessment:

1. Specifying the dimensions of school readiness
2. Establishing the criteria for choice of measures
3. Determining the capacity for conducting the assessments
4. Establishing the study population
5. Measuring error and sample-size trade-off
6. Calibrating across measurement instruments
7. Conducting matrix sampling
8. Identifying local options for additional measures or increasing sample sizes
9. Determining human participant procedures
10. Planning for secondary purposes

In the paper, I hope to focus on each of these in turn to stimulate discussion about designing readiness assessment systems. However, in the design phase, it is often necessary to circle back to ensure that in addressing a later issue, the previous decisions have not been undone.

### *1. Specifying the dimensions of school readiness*

School readiness is a composite of many attributes as Shepard et al. (1998) emphasize in their principles and recommendations to the National Education Goals Panel. Of course, readiness includes academically oriented measures, such as cognitive abilities and language development, and also indicators of social, emotional, and physical well-being. The specific list of instruments, secondary data sources, and methods of administration will not be addressed in this paper. During the design phase, when precisely specifying the dimensions to be included in the assessment of school readiness, it is important to decide whether an overall index of "readiness" will be needed.

If one of the overarching purposes is to inform the public and policymakers about the results of the cumulative investments in children, dramatic and easily understood measures will capture the greatest attention (Kingdon, 1995; Hilgartner & Bosk, 1988). Collapsing the measures into a single index provides an overall picture and avoids the "on the one hand, on the other hand" equivocation. Culturally, Americans, who are busy with making a living and their own interests, are prone to want simple score cards and ratings that capture their attention and don't require much time to understand. A single, bold assessment of school readiness may capture more public interest, and much research supports the idea that what the public considers important receives attention on the public agenda (Monroe, 1998; Page & Shapiro, 1993). The design phase is the time for reconciling the need for a broad range of indicators to help experts understand school readiness and the need for simplicity and clarity in reporting, if the issue is to receive some of the "surplus compassion" of the public (Hilgartner & Bosk, 1988). However, some states choose to report results for statewide assessments separately, for example, reporting reading, math, science, and social studies for the third grade with no attempt to combine them. The salient point is that this aspect of reporting has many ramifications that will affect the design of the assessment.

### *2. Establishing the criteria for choice of measures*

School readiness is both a goal in and of itself and a means for furthering other goals. The lists of measures used in the North Carolina Readiness Assessment and the Georgia Early Childcare Study provide intrinsically meaningful indicators. However, the public—and especially journalists—are constantly searching for the implications of these indicators. If we improve school readiness scores, will third-grade test scores rise? Will dropout rates go down? Will we have a more educated workforce? In the selection of specific dimensions and measures, the predictive validity or ability of the measures to accurately predict future consequences for children should be considered. To the extent possible, we should be prepared to provide research-based statements about the probable consequences of low scores on the Peabody Picture Vocabulary Test or the Social Skills Rating System. For example, in the development of K–12 indicator systems, the percentage of students who are two or more years over age in the eighth grade (retained twice) is often used and justified as the best predictor of dropout. In making choices between measures within dimensions and the attention given to any single dimension, the criteria of predictive validity should be considered, along with more standard criteria of reliability and other forms of validity (Shepard et al., 1998; Gilliam & Zigler, 2000).

### *3. Determining the capacity for conducting the assessments*

It is probable that the limiting conditions for conducting readiness assessments are the capacity to train assessors, manage the logistics, implement appropriate human participant protocols, and collect valid and accurate assessments. All of these are made more difficult because of the issue of timing and naturally occurring development. If the assessments are carried out over several months, it is not clear what they represent. In Georgia, for example, over 100,000 children enroll in kindergarten in about 800 public and 200

private schools. It is nearly impossible to imagine having the capacity to directly assess the population of five-year-olds by anyone other than teachers or, if large amounts of missing data can be tolerated, parents. Developments in using teachers as raters (Meisels et al., 2001) and the use of individual sampling and matrix sampling can reduce the strain on the capacity for finding, hiring, training, and managing assessors to do the field work. However, each time a layer of sophistication is added to the design (several of which are discussed in the next sections), the capacity to manage the logistics must be carefully examined. The trade-off between comprehensiveness and obtaining assessments that are valid, accurate, and reliable must be constantly assessed and reassessed when decisions are made about managing the assessment process.

### *4. Establishing the study population*

Clearly, conducting an assessment to estimate the extent to which children are ready to be successful in school involves assessing children at the beginning of their involvement with school. Whether the best approach is to assess five-year-olds starting kindergarten or six-year-olds starting first grade could be debated. In the recommendations to the National Education Goals Panel, Shepard et al. (1998) frame the issue of determining the study population pragmatically rather than ideologically. If kindergarten is the institutional base from which the study population is selected, what should be done about the children who do not attend public kindergarten? If first grade is the institutional base for the assessment, hasn't the kindergarten experience blurred the effects of preschool and school, since kindergarten has become nearly ubiquitous?

While the assessment of kindergartners raises methodological challenges for evaluating the investment in young children within a state, the design team must rule out the beginning of kindergarten as the point at which the assessments should occur. An important advantage of this choice is to confront the dearth of data on children prior to entering public education. Interpretation of the data will be forever murky if the assessment is staged at the beginning of first grade. It may also be tempting to consider assessments that fall throughout the kindergarten year. Capacity considerations (discussed later) and supporting instructional purposes (for example, informing the kindergarten teacher about developmental objectives or providing the teacher with direct information on the development of each child in the class) may cause the consideration of delayed or phased assessments. Again, the interpretations are likely to be made more difficult and less conclusive as a result. Certainly, it can be argued that the children are no worse off during the kindergarten year or at the beginning of the first-grade year than they were at the beginning of kindergarten. The cost of clarity in interpretation, especially in providing easily digestible information to the public, may be too high.

One additional issue to consider if kindergarten is chosen is the inclusion or exclusion of repeaters. These children are not beginning their school experiences. Their inclusion would affect the interpretation of the data. Yet, it is standard practice for educational assessments during elementary school to include students who have been retained in the grade.

### *5. Measuring error and sample-size trade-off*

In selecting measures, instruments, and methods of administration, there is an inherent relationship between the reliability of the instrument as administered and the sample sizes necessary to produce the same sampling error. For example, some instruments have a standard deviation of 1. To produce an estimate of the study population's value on such an instrument, a sample size of 800 would yield a standard error of .035, which would yield a confidence interval of +/- .070, or less than a tenth of a point on the scale. For an alternative instrument that is less reliable (or is being administered under conditions that reduce its reliability) with a standard deviation of 5, a sample of 8,000 would be required

to yield the same overall standard error and confidence intervals. In other words, the relationship between sample size and standard error is non-linear as shown by this case where the standard deviation increased by only five times, but the sample sizes needed to maintain a constant standard error increased by ten times. In other words, small samples may be sufficient with highly reliable instruments, but population or large samples may be assessed with much less reliable instruments and have the same degree of precision of the estimates. Therefore, if teacher ratings are less reliable, it may be possible to compensate by obtaining ratings on a larger sample of children. It should be noted that this only applies to precision and does not include bias as a source of error.

### *6. Calibrating across measurement instruments*

In the design of readiness assessment systems, it may be useful to consider a tiered plan for the administration of instruments that would embed a tightly controlled study of a relatively small sample of children within a larger assessment of a larger sample or the entire study population (see Schweinhart's chapter in this volume for an example of a tiered approach to program evaluation). The larger assessment could use much less expensive instruments and less precise measures, which might have several positive side effects, such as acquainting teachers and parents with highly important dimensions of children's readiness. The tightly controlled study could be the source for all of the indicators of readiness that are published. The multiple sources of data on the same child may enable researchers to conduct calibration studies and test for bias in lower cost assessments by either parents or teachers. This type of design has large potential for informing teachers and parents about the most salient characteristics of child development and providing some clues as to biases that routinely occur in making judgments.

### *7. Conducting matrix sampling*

Matrix sampling is a well-established assessment procedure that randomly assigns some of the assessment items to one group of children and other items to other groups of children. All items are included and can be reported on, but no child has to respond to all of them. Matrix sampling has been recommended by the Goal 1 Early Childhood Assessment Resource Group (Shepard et al., 1998) as a safeguard against using the data to make individual decisions about children and overburdening any individual child. Frequently in K–12 assessment systems, matrix sampling has been replaced because it means that comprehensive information about specific children is not available to inform parents and teachers. While reporting information on individual children may not be the purpose driving the assessment system, it is important to understand and make conscious decisions about matrix sampling and its limitations. Of course, sampling students would have some of the same safeguards as matrix sampling, although not to the same extent, and some of the same limitations.

There is one other issue that is raised by matrix sampling. Similar to the National Assessment of Educational Progress, it might be useful to develop a procedure to categorize children's developmental progress at specific steps along the continua of readiness. NAEP uses the label *proficient*, and many states use the label *competent* in their assessments to identify students who are judged to meet established performance criteria. It may be preferable to have an aggregate indicator of readiness based on the estimate of the number of children who are "ready to be successful" in school or some other way of demarking readiness. If this type of aggregation poses too high a risk for young children that cannot be addressed by human participant protections, then imposing technical impediments, such as matrix sampling, should be considered. Moreover, matrix sampling procedures do not eliminate the possibility of estimates of the number of children entering school fully ready for success in school. As Shepard et al. (1998) note, NAEP uses a form of matrix sampling, but the NAEP assessments are uni-dimensional (reading or math, not

both). Similar estimates can be constructed from carefully designed, inter-linked matrix samples, but they are more complex when the assessments cover multiple dimensions, such as the five covered in the National Education Goals Panel report (Shepard et al., 1998). A risk-benefit discussion relative to this issue should be a part of the design process and aired with key decision-makers.

### *8. Identifying local options for additional measures or increasing sample sizes*

Assessments of readiness that are being conducted at the state level are usually concerned with accurate estimates for the entire state or perhaps for a few regions within the state. For the purpose of assessing the merit and worth of the public, private, and social investments in young children to gain leverage on the policy agenda, statewide estimates are likely to be sufficient. However, at the city and county levels, policymakers and policy advocates may need more information relevant to their specific sub-state jurisdiction. In addition, some counties or school districts want to have additional measures of school readiness for their own purposes. Once again, these requests must be evaluated using capacity and risk as a part of the decision making. Serving as many needs for information as possible provides the most stable base for maintaining the system over the years. However, inaccurate information or misuse of the data to assign children to inappropriate programs or withhold services can undermine an otherwise effective assessment system.

### *9. Determining human participant procedures*

The importance of human participant procedures and the applicability in almost all situations where risk could be involved has increased in recent years. While the concern of early childhood specialists for misuse of data has been at the forefront for many years, K–12 assessment systems are often developed and implemented without explicitly considering these procedures as potential safeguards. Independent reviews are important to ensure that human participants are adequately protected. Active, informed consent (opt-in) or passive, informed consent (opt-out) of parents may be deemed appropriate. Furthermore, the consent may constitute a binding limitation on how the data could be used. These procedures may strengthen the protections for young children who participate in the assessments.

### *10. Planning for secondary purposes*

If the only purpose for assessing school readiness is to inform judgments about the public, private, and social investments in children under age six, then utilizing traditional sampling procedures to select a subset of children and matrix sampling can substantially lower costs. These techniques may be used if program improvement purposes are being added to the primary purpose. However, if instructional purposes or informing parents about their child's development are added, then the sampling methods should be avoided. It is possible to use matrix sampling and still provide information to teachers about the development of children, in the aggregate, coming into their schools. However, the sampling procedures must then include students in every school and include them in sufficient numbers to develop a reasonably accurate estimate for the school. This will require more children—perhaps more children than are present in most schools—if matrix sampling is used.

If organizational and program improvement purposes are considered, an important factor is whether the objective is to estimate outcomes for children who have participated in a particular program or to determine need for the program's services among the population that was unserved. The first calls for a sufficient sample size of program participants and careful consideration about whether the non-participants who are included in the assessment could be a reasonable "comparison" group. Much published research relies on (Reynolds, 2000) or advocates (Gilliam & Zigler, 2000) post hoc, constructed comparison groups. However, the reasonableness of this depends entirely on the initial

differences between children who participated in the program and those who did not. For geographically limited, means-tested programs, such as the Chicago Parent Child Centers (Reynolds, 2000), comparison groups of this type are quite reasonable, although not optimal. With a statewide pre-kindergarten program, the group of children whose parents chose not to send them to pre-k but who are included in the readiness assessment may be very different and any comparisons too biased by those differences to be usable. A second purpose, program improvement, which is more akin to a needs assessment than an outcome evaluation, would not be compromised by selection bias in the same way, but it will still be very important to be able to disaggregate the readiness assessment data by those who received and those who did not receive the services. This is often more difficult than it sounds, especially with children who may have used several different types of preschool or childcare during the previous year. If it can be assembled accurately, information about program coverage could also serve oversight purposes. Knowledge-development purposes could raise any number of design issues, but they are very specific to the research question being pursued and, therefore, too numerous to provide any general guidelines in this paper.

## Conclusion

Assessment systems are costly in terms of the funds required to develop and maintain the system and the opportunity costs to children, teachers, and others. It is easy to recognize the need for assessing school readiness and to justify that the public should absorb these costs, but system designers and implementers must be careful stewards of the public's coffers. A prerequisite for developing these systems is a clear statement of purpose. This paper has detailed the need for assessing the merit and worth of private, public, and social investments in young children. This would fill a gap that currently exists between indicators that are measured at birth that relate to child development and test scores from early years of elementary school. Measures of readiness taken at entry to school can indicate the need for the types of solutions that have been shown to benefit young children. Policy research indicates that problems and solutions must travel in tandem (Kingdon, 1995) if they are to find a place on the policy agenda and be enacted. Assessments of the readiness of young children for school can answer questions about why we need to enact beneficial policies and programs and, ultimately, to contribute to the well-being of children.

It seems that human nature leads us to see the potential in things we believe to be important and to begin to add purposes to the list of things a readiness assessment system should be designed to do. Potential is a great burden. All too often, trying to serve too many purposes leads to serving none. However, not meeting purposes that are highly valued by the public, professionals, or policymakers can derail assessment systems. In the K–12 arena, many norm-referenced tests have been replaced because they do not provide information specific to a state's standards. In addition, assessments that relied on matrix samples have been replaced because they do not provide information on individual students for teachers and parents. Obtaining input from the public and policymakers in ways in which they can understand the trade-off implied by adding a purpose or eliminating one, and doing so in advance, can reduce second guessing after the information is reported and it becomes evident that all potential purposes have not been fulfilled. Systematic evidence about the most valued aspects of an assessment system (Mark, Henry, & Julnes, 2000) may be used to obtain guidance on the most highly valued purposes and high-priority outcomes as well. If an inquiry about the value of various purposes is attempted, it is extremely important to provide descriptions that include cost, testing time, and information use for likely alternative designs to get informed opinions. This may add time to the design process, but in the contentious halls of modern democracies, spending time to do this may contribute to the development of a readiness assessment system that lasts.

## References

Barnett, W. S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *The Future of Children, 53*(3), 25–49.

Bloom, H. S. (1999). *Estimating program impacts on student achievement using "short" interrupted time series.* The Manpower Demonstration Research Corporation.

Cauthen, N. K., Knitzer, J., & Ripple, C. H. (2000). *Map and track: State initiatives for young children and families.* New York: Columbia University, National Center for Children in Poverty.

Gilliam, W. S., & Zigler, E. F. (2000). A critical meta-analysis of all evaluations of state-funded pre-school from 1977 to 1998: Implications for policy, service delivery and program evaluation. *Early Childhood Research Quarterly, 15*(4), 441–472.

Harkreader, S. A., & Henry, G. T. (2000). Using performance measurement systems for assessing the merit and worth of reforms. *American Journal of Evaluation, 21*(2).

Henry, G. T., & Rubenstein, R. (in press). Paying for grades: Impacts of merit-based financial incentives on educational quality. *Journal of Policy Analysis and Management.*

Hilgartner, S., & Bosk, C. (1988). The rise and fall of social problems: A public arenas model. *American Journal of Sociology, 94*(1), 53–78.

Horm-Wingerd, D. M., Winter, P. C., & Plocfchan, P. (2000). *Primary level assessment for IASA Title I: A call for discussion.* Washington: Council of Chief State School Officers.

Katz, L. G. (1997). *A developmental approach to assessment of young children.* Champaign, IL: University of Illinois.

Kingdon, J. (1995). *Agendas, alternatives, and public policies* (2nd ed.). New York: Harper Collins.

Mark, M. M., Henry, G. T., & Julnes, G. (2000). *An integrated framework for understanding, guiding, and improving policies and programs.* San Francisco: Jossey-Bass.

Meisels, S. J., Bickel, D. D., Nicholson, J., Xue, Y., & Atkins-Burnett, S. (2001). Trusting teachers' judgment: A validity study of a curriculum-embedded p. rformance assessment in kindergarten to grade 3. *American Educational Research Journal, 38*(1), 73–95.

Mintrom, M. (1997). Policy entrepreneurs and the diffusion of innovation. *American Journal of Political Science, 41*(3), 738–770.

Molnar, A., Smith, P., Zahorik, J., Plamer, A., Halbach, A., & Ehrle, K. (1999). Evaluating the SAGE program: A pilot program in targeted pupil-teacher reduction in Wisconsin. *Education Evaluation and Policy Analysis, 2*(2), 165–177.

Monroe, A. (1998). Public opinion and public policy: 1980–1993. *Public Opinion Quarterly, 62*(1), 6–27.

Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *Future of Children, 5*(2), 113–127.

Page, B. I., & Shapiro, R. Y. (1992). *The rational public: Fifty years of trends in Americans policy preferences.* Chicago: University of Chicago Press.

Peisner-Feinberg, E. S., & Burchinal, M. R. (1997). Relations between pre-school children's child-care experiences and concurrent development: The cost, quality, and outcomes study. *Merrill-Palmer Quarterly, 43*(3), 451–477.

Reynolds, A. J. (2000). *Success in early intervention: The Chicago child-parent centers.* Lincoln, NE: University Nebraska Press.

Reynolds, A. J., Temple, J. A., Robertson, D. L., & Mann, E. A. (2001). Long-term effects of an early childhood intervention on educational achievement and juvenile arrest. *Journal of the American Medical Association, 285*, 2339–2346.

Schweinhart, L. J., & Weikart, D. P. (1997). The high/scope pre-school curriculum comparison study through age 23. *Early Childhood Research Quarterly, 12*, 117–143.

Shepard, L., Kagan, S. L., & Wurtz, E. (1998). *Principles and recommendations for early childhood assessments.* Washington DC: The National Education Goals Panel.

# *Issues in Implementing a State Preschool Program Evaluation in Michigan*

**Lawrence J. Schweinhart**, High/Scope Educational Research Foundation

## Abstract

What are the challenges of state preschool evaluations? Lawrence Schweinhart addresses design and implementation issues encountered by the High/Scope Educational Research Foundation's evaluation of the Michigan School Readiness Program. Presenting a range of practical issues, such as the cost of program evaluations, through more technically complex issues, such as validity and reliability associated with using teacher observation data, the paper shares program evaluation strategies that have been used in Michigan. Michigan's tiered program evaluation strategy—with intensive data collected from children in a select group of programs, program-quality data and child-risk-factor data collected from all programs, and support for local evaluations—is described.

## Introduction

This paper is organized as a series of questions and answers on how to design and implement an evaluation of a state preschool program. It is based on the author's experience in directing High/Scope's multifaceted evaluation of the Michigan School Readiness Program (MSRP) and providing consultation for similar efforts. Working with the Michigan Department of Education (MDE), High/Scope Educational Research Foundation has been conducting an evaluation of MSRP since 1996. The program itself began at a few sites in 1985 and expanded statewide beginning in 1988. In 2001, we began an evaluation of a state funding initiative to expand part-day MSRP and Head Start programs to full day. This is one of the only proactive, fully publicly funded full-day early childhood program efforts in the U.S.

### *How can a state preschool program evaluation establish its internal and external validity?*

The scientific worth of an evaluation depends on its internal and external validity. Internal validity builds on strong design and objective assessment to produce results that really mean what they seem to mean. External validity builds on solid sampling of programs, even participation of all local programs, to produce results that generalize or apply to all local programs in the state program.

A good way to deal with both internal and external validity is to have a multi-faceted evaluation that includes at least three components: (a) an intensive scientific evaluation, (b) extensive statewide collection of some data, and (c) technical support to grantees for local evaluations.

### *What is needed for an intensive scientific evaluation?*

An intensive scientific evaluation should involve a comparison group and observational data collected by trained data collectors. It should compare a group of children who experience the preschool program to a comparison group of children like them who did not experience any preschool program. While an experimental design involving random assignment of children to program and comparison groups is ideal, quasi-experimental designs are probably more feasible. One practical approach, which we used in the Michigan evaluation, is to select comparison children at kindergarten entry who have the same background characteristics as the program group.

Investment in the collection of observational data by trained data collectors strengthens the evaluation's objectivity and ecological validity. Although teachers collect valid observational data (Schweinhart, Oden, & Jurkiewicz, 2000), their self-interest in evaluation of their own programs creates an apparent conflict of interest for them as data collectors. For young children, observational data has much stronger ecological validity than do tests. Michigan early-childhood-assessment guidelines reflect this priority, and our Michigan evaluation relies on observational data to the exclusion of tests in the preschool year. But tests, administered one-on-one by trained testers, do provide a precise objectivity lacking in observational data, and we have used them in other early childhood program evaluations.

The Michigan evaluation has tracked one cohort of children at half a dozen sites through fourth grade this year. During the 2000–2001 program year, we began collecting data on a new cohort in the children's preschool year. The new cohort has only two sites, one urban and one rural. The intensive evaluation has found evidence of program effects on children's development, school readiness, and subsequent school success. In particular, it found that only 11% of the program group was held back a grade as compared to 21% of the comparison group (Xiang & Schweinhart, 2001).

### *What data should be collected statewide?*

The extensive statewide data collection should emulate the intensive scientific evaluation but must rely on what opportunities present themselves. It is desirable to collect statewide data on child and family background, program characteristics, and child outcomes. Child and family background is easiest to collect because some such information is required for enrollment (followed by program characteristics and, finally, child outcomes).

In Michigan, we now annually collate two types of data statewide—on children's risk factors and on program quality. To qualify for the program, children must have 2 of 25 risk factors, designated and defined by MDE. Local program staff record all of children's risk factors on optical mark forms and send them to High/Scope for collation by means of a scanner that reads optical marks. (Optical mark-reading technology is critical to inexpensive collation of statewide data.) Such data in 1998 indicated that 67% of MSRP children lived in low-income families, 42% lived in single-parent families, and 31% lived in families with a history of academic failure (Xiang, Schweinhart, Hohmann, Smith, Storer, & Oden, 2000).

In addition, local program staff assess their own programs using the High/Scope Program Quality Assessment. High/Scope developed this 72-item tool using the state standards, Head Start standards, and our own previous work. It has evidence of its reliability and validity. Staff either complete the items themselves or have others, such as administrators or fellow teachers, complete the items for them. They record the ratings on optical mark forms and send them to High/Scope for collation. From this base, High/Scope staff and MDE staff have worked closely to examine and revise the state's reporting schedule to obtain information on variables important to understanding the program. Such data in 1998 indicated high PQA scores in general (averaging 4.18 on a 5-point scale), but low

scores (of 3 or less) for significant percentages of respondents on certain items, such as professional organization affiliation (43%), time for child recall of activities (38%), and anecdotal note taking by staff (31%).

Although we do not yet collect statewide data on child outcomes in Michigan, we do have local program staff report on their local evaluations. Despite some improvement in recent years, only a minority of local grantees are reporting on child outcomes at all (Xiang et al., 2000), so we provide technical support to local grantees regarding child outcomes assessment and program evaluation.

### *What technical support for local evaluation should be provided to grantees?*

A state preschool program evaluation needs to address the issue of providing technical support to grantees for local evaluations. As I said earlier, local evaluations have been a requirement of the Michigan School Readiness Program from the beginning, but the requirement was not enforced because it was obvious that programs could not meet it without technical support and MDE was not in a position to provide it. Three years ago, the W. K. Kellogg Foundation in Michigan provided us with a three-year grant to provide evaluation assistance to MSRP grantees. The first year, we offered three types of workshops—a two-day workshop on observational assessment of children's development, a two-day workshop on program quality assessment, and a one-day workshop on evaluation design. In the second year, we continued to offer variations of these workshops. In the third year, we focused more on providing consultation to local grantees in support of their evaluation efforts.

### *What should the training for assessment be? When and how is it best done?*

Whether teachers or data collectors collect the data, they need to receive training in how to do it properly. The state and local MSRP evaluations involve data collected by data collectors and teachers. The principal data collection instruments are the High/Scope Program Quality Assessment (PQA) (High/Scope, 1998) and the High/Scope Child Observation Record (COR) (High/Scope, 1992). We provide two days of training to data collectors or teachers for either of these tools. The training involves practice with each process involved in data collection: observation, anecdotal note taking, identifying the items contained within the anecdote, and assigning the anecdote a level on each relevant item. We train data collectors to a standard of inter-observer agreement that we assess at the end of the training. We have not assessed the inter-observer agreement of teachers after training, except as part of designated studies, although a case can be made for doing so. We hire and pay data collectors.

Teachers participate in training as grantees choose, with encouragement from High/Scope and MDE. We recommend that teachers receive training prior to using an instrument, but practical considerations come into play. For example, MDE required teachers to provide self-assessments on the PQA for two years before the W. K. Kellogg Foundation gave us the funding to offer PQA training at minimal cost. Although we recommend two days of training for the PQA and two days of training for the COR, teachers sometimes use either instrument with less than the recommended training. Some teachers learn how to use the instrument from the manual, a training CD-ROM, or other materials, and other teachers learn unassisted or with the assistance of a supervisor or other user.

### *Who should collect the data and what factors should be considered in making this decision?*

In local evaluations, teachers are usually the data collectors because data collection can be built into their jobs, whereas trained data collectors are an extra program expense. Also, there is evidence that teachers can collect data that are more reliable and valid than the data collected by data collectors (Schweinhart et al., 2000). Teachers may have a self-interest as

well as competing responsibilities, but their access to children's behavior far exceeds the access of data collectors. With the COR, for example, teachers collect data from children over six weeks or more, while data collectors usually observe four children over three part-day sessions. Further, teachers can notice that certain data are missing and make mid-course corrections to obtain complete data, whereas data collectors' missing data at the end of the three part-day sessions must remain missing data.

### *How should sites be included in decision making regarding the assessment and the data collection?*

As with any project involving people, it is desirable to include program staff in decisions about the evaluation design and instrumentation. But once set, it is critical that everyone participates in the evaluation design and instrumentation. In Michigan, the intensive scientific evaluation had an advisory panel of MSRP staff from across the state who provided input and feedback on the design and instrumentation of the study. Once the design and instrumentation of such studies are established, however, participating sites must conform to the established design and instrumentation or cease to be a part of the evaluation. In fact, one site did not participate in the initial child outcomes evaluation because it insisted on using an instrument other than the COR, thereby eliminating the site from the child outcomes analysis. These same procedures apply to the full-day evaluation, except that the participating sites themselves are serving as a de facto advisory panel.

In the statewide data collection, MDE requires sites to report data on risk factors and program quality with designated instrumentation. Local staff continues to have considerable discretion in the instrumentation and design of local evaluations. In the 1999-2000 program year, only 44% of the sites even mentioned outcomes as part of their local evaluation reports to the state; however, this was up from 25% two years earlier.

### *What is the nature of public engagement, and what should it be?*

The public needs to know more about the value of high-quality preschool programs for children at risk of school failure. Even policymakers and school administrators are not as informed as they ought to be in this regard. To the extent that the evaluation identifies benefits of MSRP, we are committed to bringing these benefits and the program itself to the attention of the public, in general, and policymakers and school administrators, in particular. To the extent that the evaluation fails to identify program benefits or identifies areas of program weakness, we are committed to bringing these findings to the attention of program staff who can take steps to surmount the problems.

The evaluation has generated three annual reports that have served as the basis of public dissemination of the evaluation's methods and findings (Florian, Schweinhart, & Epstein, 1997; Xiang & Schweinhart, 2001; Xiang et al., 2000). Each annual report is accompanied by a brief executive summary or fact sheet for broader dissemination. We have made these materials available to local grantees and posted them on the Web. We have also targeted specific audiences for dissemination through their conferences or other meetings. Dissemination to MSRP and Head Start audiences has been well received, but dissemination to educators, policymakers, and the public through reporters has received mixed results. On the other hand, a little bit of success in such dissemination seems to go a long way. For example, for the last annual report, we worked with MDE's public information officer to develop and disseminate a press release. In the course of doing so, he had to get it approved by the State Superintendent and send it to the Governor's Office to see if Governor Engler wanted to be involved in releasing it. Although the Governor did not pursue further involvement, the effort drew his and the Superintendent's attention to the evaluation's positive findings. Not much press appeared to result from the effort, but these other valuable communications did take place.

Our communication with legislators has not been as direct as we had expected it would be. We thought we would be asked to testify for various legislative committees regarding annual appropriations for the program. That did not happen, but it seems likely that the annual reports and knowledge among policymakers of the positive evaluation have contributed to the steady increases in funding for the program. Indeed, from 1988–1989 to 2000–2001, part-day MSRP funding increased almost six fold, from $15 million to $85.5 million. The evaluation surely did not cause this increase, but it did encourage it, whereas a negative evaluation would have raised a serious question about such expansion.

***What are the costs of creating assessment systems? What are the comparative costs of one strategy over another? What is the most cost-efficient way to design and administer systems?***

MDE has provided High/Scope $300,000 a year for the intensive scientific evaluation and statewide data collation and $225–240,000 a year for the full-day evaluation. The W. K. Kellogg Foundation provided $150,000 a year for the evaluation support to local grantees, but this level of funding met only part of the need. On the other hand, providing training in evaluation underscored the dearth of training in curriculum and program operation.

These are only the visible costs of such evaluation efforts. Evaluation is intrinsic to program operation, so many of its costs are embedded within program operation costs. The success of our local grantee evaluation support effort was no doubt conditioned on the program requirement that MSRP teachers be certified teachers with an early childhood endorsement. By the same token, the success of this effort was relative and clearly could have been greater. Although outcomes evaluation is now more common in local evaluations, it is not yet universally required as it ought to be.

Such observations should be put in perspective, however. Preschool programs for at-risk children have much in common with kindergarten programs, both in the similar ages of children and in the fact that both were introduced into school systems that did not originally include them. Froebel invented the kindergarten in 1840 with religious and philosophical justification, and today kindergarten is nearly universal in the U.S. In all those years, so far as I know, not a single experimental study has ever demonstrated the value of kindergarten. Some years ago, we worked with the South Carolina Department of Education to find that children who went to kindergarten had higher first-grade test scores than those who did not (Schweinhart & Barnett, 1984). However, this finding could have been due solely to better-off children going to kindergarten rather than kindergarten's contribution to their development. Some studies have found mixed evidence that full-day kindergarten programs have stronger results than part-day programs (Rothenberg, 1995); these studies only underscore the lack of any studies establishing the worth of part-day kindergarten programs in the first place.

We must conclude that our zeal for empirically proving the worth of preschool programs has taught policymakers to insist on evaluations that prove their worth. Without our efforts, policymakers might well accept preschool programs on their apparent merits without insistence on rigorous evidence of these merits. Ultimately, our efforts and their response grow from a cultural worry about whether organized educational programs are good for young children. Our evidence feeds this worry by showing that some programs contribute to children's development while others do not. So, perhaps evaluation and preschool programs are meant for each other. If they are, though, evaluation is also meant for programs at kindergarten, elementary school, middle school, high school, and postsecondary institutions.

## References

Florian, J. E., Schweinhart, L. J., & Epstein, A. S. (1997). *Early returns: First year report of the Michigan School Readiness Program evaluation.* Ypsilanti, MI: High/Scope Educational Research Foundation.

High/Scope Educational Research Foundation. (1992). *The High/Scope Child Observation Record for ages 2½ to 6.* Ypsilanti, MI: High/Scope Press.

High/Scope Educational Research Foundation. (1998). *The High/Scope Program Quality Assessment: Pre-school version.* Ypsilanti, MI: High/Scope Press.

Rothenberg, D. (1995, May). Full-day kindergarten programs. *ERIC Digest*, EDO-PS-95-4. Retrieved from **www.ericeece.org**

Schweinhart, L. J., & Barnett, W. S. (1984). *South Carolina's early childhood programs are a good investment* (Report to the Office of the Governor of South Carolina). Ypsilanti, MI: High/Scope Educational Research Foundation, ERIC Document No. ED 252 273.

Schweinhart, L. J., Oden, S., & Jurkiewicz, T. (2000, June 28). *Do teachers or observers see children's development better?* Poster presented as part of the Head Start Quality Research Consortium Symposium at Head Start's Fifth National Research Conference, Washington, DC

Xiang, Z., & Schweinhart, L. J. (2001). *Ready for success: Annual report of the Michigan School Readiness Program longitudinal evaluation.* Ypsilanti, MI: High/Scope Educational Research Foundation. Retrieved from **www.highscope.org**

Xiang, Z., Schweinhart, L. J., Hohmann, C., Smith, C., Storer, E., & Oden, S. (2000, February 28). *Points of light: Third year report of the Michigan School Readiness Program evaluation.* Ypsilanti, MI: High/Scope Educational Research Foundation. Retrieved from **www.highscope.org**

# *Instrumentation for State Readiness Assessment:*

## Issues in Measuring Children's Early Development and Learning

**John M. Love**, Mathematica Policy Research

## Abstract

How can we measure children's readiness for success in school? This paper reviews important elements of what we are trying to measure (elements of readiness as a concept), the context within which we are measuring children's readiness for success in school, and criteria for selecting instruments. The paper concludes by pointing out that readiness is a broad concept and in order to adequately assess children's readiness for success in school, one must take into consideration the vast array of children's experiences that have culminated in their "readiness" as they enter school.

## Introduction

State systems for assessing children as they enter kindergarten are expanding rapidly. Ever since the nation's governors created the National Education Goals Panel some 11 years ago, the first goal, that "By the year 2000, all children in America will start school ready to learn," has stimulated widespread discussion and debate and, even more, has led to action on many fronts. It seems as though each week sees more states entering the assessment arena, intent on determining whether their children are "ready to learn" when they come to kindergarten. Many of us are called on for advice by state administrators, school district evaluation staff, or funders. We get phone calls and e-mails with the plea: "We're under the gun to produce some scores; what should we do? And by the way, we have to collect data in September" (the request having come in April or May, if that early). Among all the issues of design, implementation, and instrumentation that surround state decisions about readiness, instrumentation often includes the thorniest problems and is often the first concern we hear about. So, what should we do, and where should we start?

I have to admit that I generally try to avoid giving concrete advice. As you will soon see, I continue that tradition in this paper. I have several reasons for taking this stance. Perhaps if I share them with you, we can, together, get a clearer picture of the substantial challenges facing those on the front lines who really must decide on the instruments and put some assessment procedures in place by Labor Day. First, however, I would like to return to the conceptual foundations of the assessment and instrumentation dilemmas. Thus, in the first section of this paper I review what seem to be the most essential ingredients of a concept of readiness. Then, I want to consider the current educational and policy context, which differs in some important ways from the events swirling around the Goal One committee in the


BEST COPY AVAILABLE

era that began these debates. A major difference is the contemporary emphasis on reading in the primary grades, and in this context, I want to remember the "basics." In the third section of the paper, I address the most central dilemmas of instrumentation—what can we know about children's readiness, and how can we choose among available instruments? Finally, I turn to a broader framework, one that I encourage us all to think about as we go about our tasks of designing, measuring, and interpreting school readiness assessments.

## What Is Readiness?

**Readiness**

- *Has five dimensions*
- *Depends on supporting conditions*
- *Is a relationship between the child's characteristics, the supporting conditions, and the nature and expectations of the school*

The Goal One Technical Planning Group broke new ground by defining not only what the important dimensions of "readiness" are but also what conditions are critical for supporting those dimensions (Kagan, Moore, & Bredekamp, 1995). The five dimensions of early development and learning (physical and motor, social and emotional, approaches toward learning, language development, and cognition and general knowledge) have become widely accepted, in one form or another.[1] Please see Attachment 1 for a summary of the five dimensions. The three supporting conditions (having access to quality preschool programs, parents as children's first teachers, and appropriate nutrition and health care) have even been expanded upon by others in recent years.

Sam Meisels (1998) notes, as have others, that the term "readiness" describes a relationship rather than a particular quality or set of characteristics of the child. In other words, if two children have the same set of developmental skills, abilities, and attitudes, one could be considered "ready" for school and the other not, depending on the nature and expectations of the school that the child will be entering. This relativity becomes even more complex when we consider that different states, and perhaps counties and communities within states, may have different expectations. In fact, I have argued previously that the community context for readiness is a particularly important consideration.

At some point, however, principle must give way to practicality. It is important to remember this relativity, but if we dwell on it, we will never move forward. I resolve this dilemma for myself by assuming that schools, at least in some general sense, are likely to have common expectations for the children who enter their kindergartens and first grades. It would be important, however, to determine whether this assumption has any validity. I return to the interesting notion of the community context of readiness in a couple of pages.

## The Current Policy Context

With the change in administrations, we have seen an increasing emphasis at the federal level on reading as the central challenge of elementary schools, and preparing children to learn to read as the major goal of kindergartens and programs that precede kindergarten—especially the year or two immediately preceding kindergarten, which we usually call "preschool." I embrace these emphases but suggest here that the early childhood field has yet to fully realize their implications for pre-kindergarten education, the myriad programs that preschool-age children experience in this country, and the assessment of school readiness.

### *The Centrality of Reading*

Schools have always focused on reading instruction in the early elementary years. Today, that focus appears even greater, and concerns about "pre-reading" extend the discussion to preschool and earlier. The current emphases for elementary school curricula have important implications for readiness assessment.

At the July 2001 White House Summit on Early Childhood Cognitive Development, Secretary Paige noted the possible chain of events set off by children who can't read:

They can't do homework, have difficulty keeping up in other classes, are repeating grades, get misidentified as learning disabled, and are shunted to special education classes. Often these kids ultimately drop out. Tommy Thompson added another dimension: "without reading skills, you can't figure out a medical prescription, read a warning label, or keep up with news that could benefit your health." Russ Whitehurst, Assistant Secretary for Educational Research and Improvement, emphasized the role of "pre-reading skills," while giving us a heads-up on assessment priorities: "Given the strong predictive relationship between pre-reading skills and later reading outcomes, screening children for pre-reading knowledge should become as routine as screening for problems in hearing and vision." To Whitehurst, pre-reading skills include "the skills, knowledge, and attitudes that are precursors to children's ability to read and write, and the environments that support those abilities." This does not sound very far from the position of the Goal One Technical Planning Group.

The summit speakers did more than talk about the outcomes for children, however. Susan Neuman, the then Assistant Secretary for Elementary and Secondary Education, emphasized the role of environmental stimulation. Similarly, Reid Lyon, NICHD's Child Development and Behavior Branch Chief, in summarizing the conference themes, noted, "School readiness concepts are best learned when provided in safe environments where the kids feel emotionally secure and where they can develop close relationships with other children and caring adults." He said that getting children ready to read is critical because of the strong "link between what preschool kids know about words, sounds, letters, and print, and later academic performance." However, what most intrigued me about his comments, which I did not see reflected as much in the other presentations, was his placement of academic achievements in a broader context: When he said that "providing opportunities to foster these cognitive abilities must be integrated in a seamless manner with interactions to develop social competencies and emotional health to reflect the *inseparable nature of these developmental achievements*" (emphasis added). I will return to these viewpoints in the final section of this paper, but first, I want to review the important considerations in defining readiness for success in school.

## Back to Basics: Let's Remember the Roots of Readiness

Three elements are crucial to the way I think about readiness: comprehensiveness, embeddedness, and continuity. The five dimensions referred to above (and summarized in Attachment 1) make the definition comprehensive. The direct implication for states is that their assessment strategies should tap all five dimensions of children's development and learning. This may not be as hard as it sounds. We already see examples on the federal level of large-scale studies that come very close to meeting this goal. For example, the constructs being assessed in the Family and Child Experiences Survey (FACES) align quite nicely with the dimensions outlined by Kagan et al. (1995, see Attachment 2). The Early Childhood Longitudinal Study-6 Kindergarten Cohort (ECLS-K), as well as the birth cohort study (ECLS-B), includes measures that span the five dimensions. Interestingly, the Early Head Start national evaluation, although focusing on children younger than the preschool years, also includes measures that span the five dimensions of early development and learning and the conditions that support them.[2]

**Critical elements of readiness**
- *Comprehensiveness*
- *Embeddedness*
- *Continuity*

Covering all the dimensions is a challenge. None of the studies cited capture all dimensions equally well because comprehensive assessment (a) requires extensive assessment with potential risk of overly intruding on the time of children, parents, and teachers and (b) means assessing areas for which the field often lacks reliable and valid measures. The areas of social-emotional development and approaches toward learning are

**Challenges of assessing all the readiness dimensions:**

- *Requires extensive assessment*
- *Requires assessing areas for which the field often lacks reliable and valid measures*

typically the least well measured. Nevertheless, it is a challenge that is important to meet head-on, since a partial assessment runs the risk of creating a biased—or, at best, incomplete—view of states' and communities' progress toward their readiness goals.

Assessment must also be designed to obtain data on the conditions supporting children's development described in the objectives accompanying the first goal. The FACES measures include a parent interview that taps a portion of the community conditions supporting readiness, but these need to be expanded. In fact, there is good reason to believe that the conditions supporting readiness to succeed in school extend well beyond the three areas outlined by the planning group. These conditions might include such elements as (a) child and family conditions (both protective and risk factors, which include child health conditions, family income, and family life conditions), (b) community service provisions and their accessibility (including health, parenting education, childcare and early education services, and the "readiness" of the schools), and (c) systems capacity (such as the efficacy and efficiency with which the community infrastructure functions).[3]

Thus, we see readiness assessments as embedded in these supporting conditions. Previously, I have argued for a community-oriented perspective to best reflect this "embeddedness" (Love, Aber, & Brooks-Gunn, 1994). The strength of a community-based approach to defining and assessing children's level of preparation for success in school is that the community can ascertain—and influence—the measures of success employed by the schools. When considering statewide assessments, we come to an interesting issue of what defines the "community." Can an entire state constitute a "community," or must we consider smaller, more-homogeneous subdivisions?

Child and family conditions impinge directly on each child's development. To further illustrate what may be important for a comprehensive readiness assessment, Larry Aber, Jeanne Brooks-Gunn, and I suggested that assessments should include indictors of the extent to which the family is "thriving, safe, or in danger across a number of dimensions of well-being" (Love et al., 1994). As with the dimensions of children's development and learning, the supporting conditions may also be tailored to what's important in each "community." They may include, for example, (a) fewer families living in unsafe housing or violence-prone neighborhoods, (b) reduced incidence of child abuse and neglect, (c) increased parental confidence that their children have a bright future, (d) increased involvement of fathers in the lives of their children, and (e) healthy marriages.

The third element is continuity. Here, I intend to characterize the nature of children's experiences leading to school entry. This might be considered to be another element of the supporting conditions. It seems worth highlighting, however, since it refers to relationships among a potentially large number of supporting conditions that can extend over a number of years. The first five years of life can lend stability to the child's development through continuity of experiences with family and programs, including childcare, Early Head Start, Head Start, pre-kindergarten programs, and various services, or the period can be disruptive to healthy development. Children who experience childcare and other out-of-home care and education settings are at some risk of bouncing from one type of program to another. Programs can provide continuity through their service emphases, or they can be so different as to cause disruption as children move from one setting to another. However, if we want to understand how well children are prepared for school—and *why*—then readiness assessments should include some measure of the range of children's program experiences over time, as well as the continuity of that experience from birth to school entry.

## Instruments for Readiness Assessment

The assessment issues related to instrumentation center around the age-old issue: "What can we know and when can we know it?" To answer that question, we need to look at advances in early childhood assessment over the last decade and think about the tough decisions that have to be made.

### *What Can We Know, and When Can We Know It?*

The early childhood field has made tremendous strides in the science of measuring important aspects of children's early development and learning. As I've already noted, a number of national studies—including FACES, ECLS-K, and, for younger children, the Early Head Start program evaluation—have measured developmental constructs and supporting conditions not heretofore included in large-scale assessments. In fact, whereas two decades ago, many of us in this field despaired of ever having a sufficient number of good-quality instruments, the concern now is a glut of instruments, along with the challenging task of selecting among them.

SERVE, for example, has recently published a compendium of assessment instruments (Niemeyer & Scott-Little, 2001). The compendium summarizes 39 commercially available instruments and, in an easy-to-use format, guides potential users to key features of each instrument that should be considered in selecting the instruments to use. Child Trends, working with Lisbeth Schorr and the Pathways Mapping Project, has compiled a compendium of measures used in national, state, and local data collections (Calkins, Ling, Moore, Halle, Hair, Moore, & Zaslow, 2001). It includes largely unpublished sources, while taking a different approach by providing item-level information. This compendium has the advantage of summarizing measures that have been used in data collections that may parallel data collections that states may want to launch for readiness assessment. As with the SERVE compilation, it evaluates the measures along a number of important criteria. No longer can we complain that measures either do not exist or are impossible to access. The challenge instead is one of sorting and sifting, to choose what will be most appropriate and useful for each state's or community's purpose.

Unfortunately, the extensive work already done for us, as reflected just in these two compendia, gets us only to the starting gate. Each of us who wants to obtain valid and reliable data on important dimensions of children's early development and learning must sift through the hundreds of measures and apply some rather complex, and often conflicting, criteria.

### *Why the Choices Are So Difficult: There Is No Perfect Answer*

It is almost impossible to list all the reasons why there is no perfect answer. As we all know, each measure has strengths and weaknesses, so, even listing the important criteria for evaluating available measures merely highlights the challenges. Nevertheless, the challenges are important, and we must meet them. Even though these criteria have been listed in many places, it may be useful to attempt a consolidated list. Some criteria apply to individual measures, but additional factors must be considered when weighing the appropriateness of the collection of measures that will comprise the "readiness" assessment. Here are what I consider the most important criteria to be.

#### Criteria for Choosing Individual Measures

1. Does it measure what it claims to measure?
2. Does it do so reliably?
3. Has it measured what you want it to measure reliably and validly, under field conditions similar to those where you will use it?
4. Does it tap an important dimension of children's early development and learning or of the conditions supporting development?

5. Is the measure appropriate for the diversity of children in your state or community, including considerations of socioeconomic status, geographic regions, racial/ethnic background, linguistic groups, and disability status?
6. Is it appropriate for the age(s) of the children you are interested in?
7. Is it available to you at the time you need it?

**Criteria for Evaluating the Final Set of Measures**

1. Does the set encompass all of the dimensions of children's early development and learning?
2. Does the set of measures also span the conditions supporting early development *and learning that are important for your locale?*
3. Will analysis of the measures provide aggregate data that will allow you to focus on the collective status of entering kindergartners?
4. Does the collection of measures incorporate multiple modes of assessment (such as direct assessment, parent or teacher ratings, observations, and self-report) so that the final judgment about "readiness" does not hinge on just one or two methods?
5. Are multiple perspectives included, such that ratings do not reflect only teacher judgments but those, for example, of the parents as well?
6. Do the measures, overall, provide a balance of positive and negative indicators of development and learning?
7. Is it feasible (meaning, also affordable) for the data to be collected with quality and consistency across the varied settings in which the assessments need to be completed?
8. If the answer to number 7 is uncertain, can the set of measures be adapted to local circumstances while retaining their essential ingredients?
9. Do some of the measures allow you to compare results with national data?[4]

Consideration should also be given to the process of preparing for any large-scale assessment. Who participates in the decision making and how the participants are involved are important considerations, as are the issues of design and implementation. Although the process will differ with each state and community, it is important that there be open discussion about the process. In many respects, readiness assessment meets all the criteria for "high-stakes" testing, which engenders so much concern and controversy among the many stakeholders in our children's educational futures.

## "Early to Learn": Developing a Broader Perspective

*Kindergarten represents the culmination of a child's first five years of life.*

In the concluding section of the paper, I want to briefly suggest a perspective with which to view the assessment enterprise. In a number of respects, readiness assessment—or, as I prefer to think of it, assessment of children's early development and learning upon entry into school—can be thought of as program evaluation but evaluation in a very special sense. Although kindergarten entry is the beginning of a long educational journey for each child, it also represents the culmination of his or her first five years of life. What children know, what they can do, what attitudes and inclinations they have—all are a function of the families they have lived in, the neighborhoods in which they have played, the many (or few) caring adults who have nurtured them (or not), and the programs and activities they have participated in (or not). America's children have taken ballet lessons, put up with bullies, enjoyed *Sesame Street* and *Mister Rogers' Neighborhood*, and been subject to violence and danger. They have attended church, mosque, temple, or synagogue, or not. All of these, and the thousands of other experiences in children's early lives, contribute to

their language, their cognitive and physical abilities, their emotions and social skills, and the way they approach new learning opportunities. Our challenge is to find a practical assessment process that will capture the "outcomes" of these vast and varied experiences.

In my "theory of change" about the first five years of life and what outcomes they should lead to, the expected outcomes look very much like the dimensions of early development and learning that now stand for readiness. It is for these reasons that I put so much stress on finding the measures that do justice to the full and comprehensive dimensions of readiness and to administering them in a way that allows the results of this five-year process to be seen. Let us not shrink from moving forward to screen all children on hearing, vision, and pre-reading knowledge, while at the same time understanding the inseparable nature of the children's range of developmental achievements.

## End Notes

*1* While this committee of the National Goals Panel proudly—and for good reasons—described "children's early development and learning" without using the term "readiness," it is increasingly awkward to engage in constructive discussions about the issues without it. Discussing "readiness for success in school" seems to avoid some of the problems with traditional use of "readiness for school" or, even worse, "readiness to learn."

*2* Administration on Children, Youth, and Families. (June 2001). *Building their futures: How Early Head Start programs are enhancing the lives of infants and toddlers in low-income families.* Washington, DC: U.S. Department of Health and Human Services. Also see Chapter 2 in ACYF (December 1999). *Leading the way: Characteristics and early experiences of selected early Head Start programs; Volume I: Cross-site perspectives.* Washington, DC: U.S. Department of Health and Human Services.

*3* Although systems capacity is a very important consideration, its assessment probably is beyond the scope of most community readiness efforts.

*4* I do not think this is an essential criterion; nevertheless, it should be considered as a possible advantage for interpreting assessment results.

## References

Calkins, J., Ling, T., Moore, E., Halle, T., Hair, B., Moore, K., & Zaslow, M. (2001). *School readiness indicator items.* Washington, DC: Child Trends, Inc.

Kagan, S. L., Moore, E., & Bredekamp, S. (1995). *Reconsidering children's early development and learning: Toward common views and vocabulary.* Washington, DC: National Education Goals Panel.

Love, J. M., Aber, J. L., & Brooks-Gunn, J. (1994). *Strategies for assessing community progress toward achieving the first national educational goal.* Princeton, NJ: Mathematica Policy Research, Inc.

Meisels, S. J. (1998). *Assessing readiness.* Ann Arbor, MI: Center for the Improvement of Early Reading Achievement.

Niemeyer, J., & Scott-Little, C. (2001). *Assessing kindergarten children: A compendium of assessment instruments.* Tallahassee, FL: SERVE.

# Attachment One

## "Readiness" Dimensions Identified by the Goal One Technical Planning Group of the National Education Goals Panel

### *A. Dimensions Based on the Major Conditions that Support Readiness*

The family and community conditions that support readiness are spelled out in the three objectives that accompany the goal itself:

- "All children will have access to high-quality and developmentally appropriate pre-school programs that help prepare children for school.
- Every parent in America will be a child's first teacher and devote time each day helping his or her pre-school child learn; parents will have access to the training and support they need.
- Children will receive the nutrition and health care needed to arrive at school with healthy minds and bodies, and the number of low-birth weight babies will be significantly reduced through enhanced prenatal health systems" (Kagan, Moore, & Bredekamp, 1995).

### *B. Dimensions and Criteria of Children's Early Learning, Development, and Abilities*

Each of the five dimensions of early learning, development, and abilities includes a number of criteria for assessment.

**1. Physical Well-Being and Motor Development**

- *Physical development* (rate of growth, physical fitness, and body physiology; prevention of diseases; disabilities)
- *Physical abilities* (gross-motor skills, fine-motor skills, sensorimotor skills, oral motor skills, and functional performance)
- *Background and contextual conditions of physical development* (the perinatal context, caregiving environment, and health care utilization; vulnerabilities, such as prenatal alcohol exposure; environmental risks, such as harmful aspects of the community environment)

**2. Social and Emotional Development**

- *Emotional development* (feeling states regarding self and others, including self-concept; emotions, such as joy, fear, anger, grief, disgust, delight, horror, shame, pride, and guilt; self-efficacy; and the ability to express feelings appropriately, including empathy and sensitivity to the feelings of others)
- *Social development* (ability to form and sustain social relationships with adults and friends, and social skills necessary to cooperate with peers; ability to form and sustain reciprocal relationships; understanding the rights of others; ability to treat others equitably and to avoid being overly submissive or directive; ability to distinguish between incidental and intentional actions; willingness to give and receive support; ability to balance one's own needs against those of others, creating

opportunities for affection and companionship; ability to solicit and listen to others' points of view; being emotionally secure with parents and teachers; being open to approaching others with expectations of positive and prosocial interactions, or trust)

**3. Approaches Toward Learning** [1]

- *Predispositions* (gender, temperament, and cultural patterns and values)
- *Learning styles* (openness to and curiosity about new tasks and challenges; initiative, task persistence, and attentiveness; approach to reflection and interpretation; capacity for invention and imagination; and cognitive approaches "styles" to tasks)

**4. Language Development**

- *Verbal language* (listening, speaking, social uses of language, vocabulary and meaning, questioning, and creative uses of language)
- *Emerging literacy* (literature awareness, print awareness [including assigning verbal labels to familiar letters, sound-letter combinations, and recognizing own name in writing], story sense [beginning, middle, end], and writing process [ordered scribbling, producing writing configurations])

**5. Cognition and General Knowledge**

- *Knowledge* (physical knowledge, logico-mathematical knowledge, and social conventional knowledge)
- *Cognitive competencies* (representational thought, problem solving, mathematical knowledge, social knowledge, and imagination)

## End Notes

1 Kagan, Moore, and Bredekamp note that approaches toward learning are particularly important for success in school because the "mere acquisition of knowledge, skills, and capacities is an insufficient criterion" without children's *inclination* to marshal these skills.

# *Attachment Two*

**How Head Start Performance Measures Are Aligned with the Goal One Dimensions**

| *Goal One Learning & Development Dimension* | *Head Start Performance Measures* |
|---|---|
| *Physical well-being and motor development* | *Gross and fine motor skills* |
| *Social and emotional development* | ➡ *Positive social behavior*<br>➡ *Personal maturity*<br>➡ *Behavior problems*<br>➡ *Social interaction with peers*<br>➡ *Social awareness*<br>➡ *Relationships with adults* |
| *Approaches toward learning* | ➡ *Creativity*<br>➡ *Initiative*<br>➡ *Attitudes toward learning*<br>➡ *Task mastery* |
| *Language development* | ➡ *Emergent literacy and language skills*<br>➡ *Receptive vocabulary*<br>➡ *Letter recognition*<br>➡ *Book knowledge*<br>➡ *Print awareness* |
| *Cognition and general knowledge* | ➡ *Numerical skills*<br>➡ *General memory*<br>➡ *Color naming*<br>➡ *Reasoning*<br>➡ *Problem solving*<br>➡ *Musical ability* |

## Measures Used in the Head Start Family and Child Experiences Survey (FACES), 1996–2001*

### Child

***Howes Peer Play Scale***
***Social Awareness Tasks***
***Peabody Picture Vocabulary Test-III***
***Phonemic Analysis Subtest (TOLD-3), K and 1***
***ECLS-K Reading and General Knowledge Assessments (K and 1)***
***Child Health Profile (First Grade)***
***McCarthy Draw-A-Design***
***Color Names and Counting***
***Woodcock-Johnson Letter-Word Identification (4 years and older)***
***Woodcock-Johnson Applied Problems (4 years and older)***
***Woodcock-Johnson Dictation (4 years and older)***
***Story and Print Concepts***
***Social Behavior Ratings (Parent, Teacher, Assessor)***
***Personal Maturity Scale (selected items)(Parent and Teacher)***
***Problem Behavior Ratings (Parent and Teacher)***
***Child Observation Record (Social Relationships, Creative Representations, and Music & Movement Subscales) (Teacher)***

### Classroom

Assessment Profile Scheduling Scale
Assessment Profile Learning Environment Scale
Early Childhood Environment Rating Scale (ECERS)
Arnett Scale of Caregiver Behavior
Counts of staff/children

### Staff Interviews and Reporting Forms

Head Start Teacher Self-Administered Survey
Kindergarten Teacher Self-Administered Survey
Other staff interviews

### Parent Interviews

Family demographics
***Child's developmental accomplishments***
Parent-child activities
Disabilities
Parent involvement and satisfaction with Head Start
***Child's behavior***
Household rules
Employment and income
Community services
Childcare
Family health care
Home safety
Home and neighborhood characteristics

Parent's feelings including:
- CES-D Depression Scale
- Pearlin Mastery Scale
- Family Support Scale

## Case Study Home Visit Parent Interviews

***Parents' description of Head Start child***
Primary reasons for enrolling child in Head Start
Hopes and goals for Head Start child
Perceptions of family strengths
Perceptions of areas for family improvements
Perceptions of family problems that may interfere with child's adjustment to Head Start
A typical day for Head Start child and family
Family's participation and satisfaction with Head Start
Parenting beliefs, hopes, goals, and satisfaction
Neighborhood characteristics
Home observations
Neighborhood observation checklist

## Case Study Monthly Telephone Contact Interviews

Household composition
***Child health***
Adult health
Childcare arrangements
Employment/economic status
Family participation in Head Start
Family contact with community agencies
Social support (intimate, informational, and instrumental)
Family resources
Psychological well-being (CES-D)
Significant family events
Head Start satisfaction
Transition to kindergarten

## Case Study Community Agency Interviews

Type of agency, agency services, agency goals and mission, and target population
Organization of service delivery and referral systems
Collaboration with Head Start
Perception of relationship with Head Start and satisfaction

* I am indebted to Louisa Tarullo, Commissioner's Office of Research and Evaluation, ACYF, for providing this listing of the FACES instruments. ***Bold Italics*** indicate the measures that provide data on child outcomes. FACES 2000, a second national cohort study, added the Leiter Sustained Attention Subtest to the child battery, substituted the ECERS-R for the ECERS, added the Assessment Profile Individualizing Scale, and added a father questionnaire.

# *Statewide School Readiness Assessments:*

## Challenges and Next Steps

**Martha Zaslow and Tamara Halle, Child Trends**

## Abstract

As states proceed with the work of designing strategies for school readiness assessment, selecting assessment instruments, implementing the assessments, and communicating results, they are encountering a series of challenges. This paper discusses these emerging challenges and offers some suggestions on steps states can take to start to address them.

## Introduction

A growing number of states are collecting and reporting on statewide assessments of school readiness (Saluja, Scott-Little, & Clifford, 2000; see also papers by Meunchow, Schweinhart, and Love in this volume on school readiness assessment). As states proceed with the work of designing strategies for assessment, selecting assessment instruments, implementing the assessments, and communicating results, they are encountering a series of challenges. These range from concerns about the adequacy of existing assessments of children's school readiness to concerns about how to communicate results effectively.

The purpose of this paper is to describe six challenges that are surfacing repeatedly across the states that are already in the process of implementing systems of school readiness assessment.[1] A careful delineation of the challenges emerging across these states can hopefully lay the groundwork for identifying solutions.

### *Challenge #1: To continue to build on clearly stated principles for school readiness assessments*

A first challenge concerns the need to continue to reference and build on statements of principle for school readiness assessment from past decades. Such statements of principle have had noteworthy effects on the way in which school readiness has been conceptualized, the selection of assessment instruments, and the ways in which assessments of school readiness have been used.

A major example of an articulation of principles shaping practice comes from the work reviewing assessment practices in the late 1980s. A survey conducted by the National Academy of Sciences and the National Association of State Boards of Education regarding testing practices for pre-kindergarten and kindergarten children found widespread use of IQ-like tests with a narrow cognitive focus being used to make decisions about placement and

retention for young children. Concerns were raised about the reliability and validity of some of the tests and about the use of the tests for placement when there were questions about the stability of children's scores at these ages. Questions were also raised about screening tests, intended to make decisions about whether further in-depth assessment was needed, instead being used without appropriate follow-up to make placement decisions.

Joint statements by professional organizations, researchers, and practitioners were made against the use of testing to deny age-eligible children entry into kindergarten. The articulation of principles was largely effective in shaping practice. A survey of states in the mid-1990s found progress away from the use of testing to determine kindergarten entry and retention and a diminution in the inappropriate use of screening instruments (Shepard, Taylor, & Kagan, 1996; see also Saluja et al., 2000, for a more recent survey of state assessment practices).

Another major example concerns the identification of children's school readiness as being multidimensional. A series of literature reviews regarding children's early development has consistently underscored the need to consider children's readiness for school as including but extending beyond early cognitive development. The National Education Goals Panel's review of the research resulted in a conceptualization of readiness in children as involving five dimensions: early literacy and cognitive development, socioemotional development, motor development and health, communicative skills, and approaches to learning (Kagan, Moore, & Bredekamp, 1995). In an important recent update, a review of the evidence on early development completed by the National Academy of Science's Committee on Integrating the Science of Early Development and reported on in the volume *Neurons to Neighborhoods* (National Research Council & Institute of Medicine, 2000) concludes also that school readiness includes but goes beyond cognitive development. This volume stresses the importance of social and emotional development in the early years for later academic adjustment and progress (see also Huffman, Mehlinger, & Kerivan, 2000, for a review focusing specifically on the socioemotional aspects of school readiness).

At the national, state, and local levels, strategies to measure children's school readiness have been built around the conceptualization of children's readiness as multidimensional. For example, in national data collection, the Early Childhood Longitudinal Study—Kindergarten Class of 1998–1999 (ECLS-K) and the Head Start Family and Child Experiences Survey (FACES) both include measures of the five dimensions of children's school readiness identified by the National Education Goals Panel's literature review. At the state level, efforts like North Carolina's state representative sample looking at the status of children in the state at kindergarten entry (as well as the status of the kindergarten classes and the schools receiving the children), chose assessment instruments to reflect these five dimensions of readiness in children (Maxwell, Bryant, Ridley, & Keyes-Elstein, 2001). Similarly, in Florida, a task force on school readiness assessment stated among its principles for assessment that the multiple dimensions of children's readiness should be recognized and assessed (Florida Partnership Board, School Readiness Performance Standards Workgroup, 2000). At the local level, work by John Love, Larry Aber, and Jeanne Brooks-Gunn (1994, 1999) provides a detailed "blueprint" for selecting measures of each of the five dimensions of children's school readiness for use in community monitoring of children's readiness.

A key challenge for the new state work on school readiness assessment is the need to continue to reference rather than lose sight of these statements of principle representing the views of major panels that included respected researchers and experienced practitioners. Referencing these statements of principle can help guard against drift back into using school readiness assessments for placement and retention purposes and drift back into the selection of assessment instruments around a narrow rather than multidimensional view of readiness. A further challenge is to continue to articulate additional guiding principles as work at the state level moves forward.

### *Challenge #2: To be explicit about the purposes of assessment*

Another challenge emerging across states is the need to be clear about the purpose states have in carrying out assessments of school readiness and the need for caution about the use of specific assessment tools for purposes *other* than the ones for which they were developed.

In response to a request by Congress, the National Education Goals Panel reviewed research and practice on the assessment of young children (NEGP, 1998a). Their review identified four distinct purposes for the assessment of young children: (1) assessment to support learning by individual children, (2) assessment for the identification of special needs in individual children, (3) assessment for program evaluation and for monitoring trends in geographical units like counties or states, and (4) "high-stakes accountability" assessment, used to make decisions about individual children, teachers, and classrooms (such as placement decisions for children and salary decisions for teachers).

The NEGP review on assessment of young children underscores the need to use the purpose of assessment as the starting point in selecting an assessment strategy. For example, if the purpose of assessment is to help teachers plan individualized instruction, then assessing every child within the classroom is an appropriate strategy; however, if the purpose is to evaluate a program that is being implemented in a school, assessing a representative sample of children across classrooms in that school may be more appropriate. Further, assessments carried out for each of the purposes have different technical requirements (for example, regarding reliability and validity as well as sampling) and also have different audiences for the information obtained (for example, teachers, parents, and children are the audience for information from assessments to benefit instruction, while the public and policymakers are the audience for information obtained from monitoring trends). Of particular importance here, the review urges great caution regarding attempts to use any strategy of assessment for a purpose other than the one for which it was intended. For example, screening instruments are appropriate as initial assessment tools for identifying children with special needs but are not likely to be appropriate as tools for program evaluation. Different tools are appropriate for different purposes.

Much discussion within and across states currently focuses on whether assessments originally designed to inform instruction should be used for other purposes, and if so, how. Specifically, the issue for many states is whether and how to use assessments that support learning for the purpose of providing a picture of how children in the state as a whole are faring as they enter kindergarten and progress through it. Assessments to support learning are intended to help shape the course of instruction for individual children by identifying what children know and can do and where they should proceed in their learning (see discussion in NEGP, 1998a). Such assessments occur on an ongoing basis throughout the academic year (rather than only at the start or end of the year) and are embedded within the content of curriculum. They can be collected through teacher observation, collection of samples of children's work, or asking questions of children or parents. The NEGP review notes the need for tools to guide observations of children's progress and for training in the use of such tools within such assessment systems. Recently, there have been important steps in the development and implementation of observational systems to support individual children's learning. For example, a number of states have implemented the Work Sampling System developed by Meisels (1999, 2000), with statewide training of teachers, and universal collection of data for children in public kindergarten and other grades.

The challenge for states interested in statewide monitoring of school readiness concerns cautions about using assessment data for purposes other than for what they were intended. In some states, the decision has been made to aggregate data from Work Sampling and other observational systems in order to monitor trends in children's readiness statewide over time. Muenchow (2001) notes another concern that can arise as states

put electronic repositories in place, and data from assessments to support instruction are entered into such "data warehouses." Even if the initial purpose of assessment was to support instruction and the state did not plan to use the data for monitoring trends, the warehousing of such data may leave open the possibility of use for this purpose.

Questions have been raised about whether data from teacher observations to support instruction have sufficient inter-rater reliability for the data to be used beyond the originally intended purpose. Concerns have also been raised about drift in ratings over time and the need to have ongoing rather than only initial teacher training on the observational systems. A fundamental issue is whether these assessments, intended to help chart the course for individual children and help inform parents of their children's progress, meet or even should meet the technical requirements for assessments that can provide a consistent portrayal across a community or state of children's performance.

The issues that have been raised by states could be addressed empirically in further work. For example, it might be fruitful to have a special study looking at how scores from measures to support instruction "map" onto those from direct assessments (such as those used in the ECLS-K and FACES). Studies could be carried out that directly examine the issues of inter-rater reliability in teacher ratings and of "drift" in reliability over time. These follow-up steps could help shed light on the appropriateness of aggregating and reporting on scores from assessments to support individual children's instruction.

### *Challenge #3: To ensure that assessments include children whose first language is not English, and to be clear about the purposes of such assessments*

A third challenge concerns school readiness assessments for children whose first language is not English. A number of assessments of children's school readiness are available in languages other than English, especially Spanish. State protocols for assessing children's readiness may call for use of assessments in the child's first language if proficiency in English is limited. While this may not be feasible in all languages, assessment in the major languages other than English represented in a geographical area may be an option.

Even if assessments are available in languages other than English, however, questions can be raised about the underlying goal of such assessment. Is the goal to:

- Get a sense of proficiency in the child's first language?
- Determine initial level of proficiency in English as children enter kindergarten?
- Determine mastery of English over time, as children are exposed to instruction and social interaction in English?
- Determine mastery in both the child's first language and English to underscore the importance of both languages to children's cognitive development and to functioning in multiple cultural contexts?

There are linkages here with the issue of the purposes of assessment. For example, if the purpose of assessment is to inform an individual child's *course of instruction*, then it may be most important to chart mastery of English over time or mastery of English as it co-occurs with continued development in the first language. In contrast, if the purpose of assessment is to *monitor trends over time in the state as a whole*, then it may be more important to assess initial level of proficiency in English in order to report on trends over time in the proportion of children who enter kindergarten needing to develop further proficiency.

Some states have taken very seriously the challenge of assessing not only children's readiness for school but also schools' readiness for children (NEGP, 1998b). This "interactionist" view of school readiness, focusing on the fit between children's

characteristics upon school entry and the characteristics and resources of the schools receiving the children, has been described in the work of Love and colleagues (1994, 1999) and Meisels (1999, 2000). This further perspective on school readiness raises the challenge of how to measure schools' readiness to receive children whose first language is not English. States may wish to articulate their goals for services to such children and to include measures reporting on the availability and use of such services.

A fruitful next step here might be to convene a working group of interested representatives from states, along with researchers and policymakers. Such a working group could examine current state practices in assessing children with a first language other than English, make recommendations about appropriate goals for assessment with these children, review available assessment instruments, and explore the concept of whether and how to measure schools' readiness for those children who are not yet proficient in English.

While we focus here on the challenge of assessment of children with special needs concerning mastery of English, a parallel set of issues could be articulated for the special needs of children with disabling conditions.

### *Challenge #4: To strengthen assessments of specific dimensions of school readiness*

A recurrent theme across states proceeding with school readiness assessments is concern about available measures for assessing the socioemotional aspects of children's school readiness. Measures of cognitive development and literacy generally appear to have better psychometric properties than measures of socioemotional development and often reflect more extensive work with national samples as part of the development of the measures.

A recently completed review of research concludes that there is better across-time prediction for measures of children's cognitive than socioemotional development from the preschool years or the kindergarten year through the second grade (La Paro & Pianta, 2000). While these patterns may indeed reflect on greater underlying continuity within the cognitive domain (a finding that would suggest greater importance in measuring cognitive development than socioemotional development at kindergarten entry), an important alternative interpretation is that the stronger cross-time prediction for cognitive measures may rest in stronger characteristics of the measures themselves.

Concerns about limitations with measures of socioemotional development go beyond research and practice in the area of school readiness. Such concerns were raised in a recent review of child outcome measures used in the research on childcare quality (Zaslow, Reidy, Moorehouse, Halle, Calkins, & Margie, 2002). Problems with the psychometric properties of measures were more often found for socioemotional than for cognitive measures. Further, this review found a lack of consensus about what aspects of socioemotional development to measure. Looking across the studies of childcare quality, studies encompassed measures addressing a wide range of constructs pertaining to socioemotional development, with little agreement across studies regarding which constructs to focus on. This seriously limits the capacity to compare findings across studies.

A further concern expressed by states is that the problems with measurement may be more acute with respect to positive than problematic aspects of socioemotional development. Indeed, much more extensive work has gone into developing measures of behavior problems than measures of positive social development.

In terms of next steps, there appears to be a need for work towards (a) identifying the key constructs in the domain of socioemotional development that are most important to measure in assessing school readiness, (b) identifying the strongest existing measures

of these constructs, and (c) where necessary, working towards the development of new measures, especially with respect to positive social development. Such efforts could be informative in substantive areas beyond the work in children's readiness for school.

### *Challenge #5: To work towards making explicit the connections between measures of children's readiness and resources allocated to young children*

One of the four purposes of assessment of young children (NEGP, 1998a) is tracking and monitoring trends in the well-being of young children in a geographical unit, like the state. If collected at a single point in time, data collected universally or for a representative sample of a state's children can help identify groups of children who are at particular risk or aspects of development that are of concern. If data on children at kindergarten entry are collected every year or at periodic intervals, these can serve as "indicators" of the well-being of young children (Phillips & Love, 1997) and provide trends pointing to changes in a favorable or unfavorable direction for children in a geographical unit over time.

As noted by Henry (2001), the use of assessments of young children for this purpose can be seen as reflecting the cumulative investments by a state or community in its young children. It can inform a legislature as to where investments are needed. When measures are collected annually or at periodic intervals over time, such data can be used to provide markers of whether changes in an expected direction are occurring in keeping with a public investment.

As one example, in the state of Rhode Island, an indicator was developed to measure the number of children in the state entering kindergarten with elevated blood lead levels (see description in Zaslow et al., 2001). In 1996, data reported in the Rhode Island *Kids Count* fact book indicated that approximately one in three children in the state were entering kindergarten with elevated blood lead levels and that this figure was higher (nearly one in two) for children in cities with the highest child poverty rates. The publication and dissemination of this indicator resulted in serious concern within the state about the lack of lead-poisoning prevention and treatment programs for young children; consequently, new programs were launched. Trends in this indicator continued to be monitored, and by the year 2000, the proportion of children statewide with elevated blood lead levels had declined from 36% to 13%.

Indicator data cannot pinpoint causality. It is only with experimental data that it would be possible to attribute a change of this kind over time, with confidence, to the intervention (rather than, for example, to an improved economy resulting in many families moving out of housing that exposes children to lead). Nevertheless, in this example, indicator data sufficed to highlight an area of concern where further investment in young children appeared warranted and to document changes over time consistent with the investment.

States are concerned that data used to track and monitor child well-being are often *assumed* to reflect the cumulative investments in young children, with *little or no direct measurement of the investments themselves*. In addition to collecting measures of child well-being over time, it may also be important to identify key markers of investments in young children.

The work of the Child Indicators Project (as summarized in Zaslow et al., 2001) and a newly launched project focusing specifically on school readiness indicators (School Readiness Indicators Project[2]) provides some guidance as to the kinds of measures that could be collected regarding state investments in early childhood development. A meaningful next step here would be dissemination of the knowledge gleaned from these projects regarding indicators not only of children's school readiness but also of investments in readiness.

### *Challenge #6: To work towards more effective communication of findings on school readiness to state legislatures and the public*

A final challenge emerging across states concerns the need for more effective communication of findings on school readiness to state legislatures and the public. More specifically, states are concerned with how to make findings useful and used while not oversimplifying or distorting the nature of the data in the process of communication.

One example of the potential challenge, and also of a solution, is illustrated by a strategy adopted by North Carolina in anticipation of the release of its report on the readiness of the state's children for kindergarten and the readiness of kindergartens for the children (Maxwell et al., 2001). The issue was the possible expectation by members of the state legislature and the public that school readiness data would be summarized in terms of a single score (as in "X% of the children in the state are ready for school"). Yet North Carolina's Ready for School Goal Team (Ready for School Goal Team, 2000) had decided that it was important for data collection to be in keeping with the perspective that school readiness is multidimensional, and the data collection focused on the five dimensions of readiness in children (as well as markers of readiness in schools).

In order to ensure that the nature of the findings would be appropriately anticipated and understood, the research members of the Ready for School Goal Team developed a mock-up of the report to share with the Governor and other key stakeholders before the data were available (Maxwell & Ridley, 2001). The mock-up provided one or two indicators for each dimension of children's readiness and the readiness of schools and illustrated how the distributions of scores would be portrayed (rather than a ready/not ready break for each measure). This mock-up was effective in introducing key members of the public and legislature to the nature of the findings that would eventually be released, and the report, when ready, was well received and understood.

A possible follow-up step here would be to prepare a working paper drawing together case studies of effective communication with the public and the media on issues concerning school readiness.

## Conclusion

The challenges facing states as they put school readiness assessment systems in place cut across (a) design issues—for example, the challenge of how to include all children, including those whose first language is not English, in the assessment process, and the challenge of designing an assessment system to reflect the readiness of schools as well as of children (b) instrumentation issues—for example, concerns about the measures of socioemotional development and concerns about using specific assessment instruments for purposes other than the ones for which they were developed; (c) implementation issues—for example, the issue of teacher training and inter-rater reliability on measures to inform children's instruction; and (d) communication issues—for example, the concern with communicating findings in a way that does not oversimplify or distort results.

It will be important to sustain a process of communication across states, to share experiences with the challenges already noted, to identify new challenges as they arise, and especially to work towards effective approaches for addressing the challenges.

## End Notes

*1* These issues were identified by participants at the symposium "Assessing the State of State Assessments." The states participating were California, Florida, Georgia, Maryland, Michigan, Missouri, North Carolina, Ohio, and South Carolina. We are indebted to the symposium organizers and participants and hope we have captured the issues that they raised accurately.

*2* This initiative, involving 16 states working to develop indicators of school readiness, is being funded by the Packard Foundation, the Ford Foundation, and the Kauffman Foundation. Leadership for this project is provided by Elizabeth Burke Bryant and Catherine Walsh of Rhode Island *Kids Count*.

## References

Florida Partnership Board, School Readiness Performance Standards Workgroup. (2000, May). *Proposal for the assessment of children's school readiness.* Unpublished manuscript.

Henry, G. T. (2001, December). *Assessing school readiness: System design framework and issues.* Paper presented at the Symposium on Assessing the State of State Assessments, Atlanta, GA.

Huffman, L. C., Mehlinger, S. L., & Kerivan, A. S. (2000). Risk factors for academic and behavioral problems at the beginning of school. In *A good beginning: Sending America's children to school with the social and emotional competence they need to succeed* (monograph). Bethesda, MD: The Child Mental Health Foundations and Agencies Network.

Kagan, S. L., Moore, E., & Bredekamp, S. (1995). *Reconsidering children's early development and learning: Toward common views and vocabulary.* Washington, DC: National Education Goals Panel, Goal 1 Technical Planning Group.

La Paro, K. M., & Pianta, R. C. (2000). Predicting children's competence in the early school years: A meta-analytic review. *Review of Education Research, 70,* 443–484.

Love, J. M., Aber, J. L., & Brooks-Gunn, J. (1994). *Strategies for assessing community progress toward achieving the first national educational goal.* Princeton, NJ: Mathematica Policy Research (MPR Reference Number 8), 113–110.

Love., J. M., Aber, J. L., & Brooks-Gunn, J. (1999). *Ready or not, here they come: Strategies for achieving school success for all Kansas City children.* Unpublished manuscript.

Maxwell, K. L., Bryant, D. M., Ridley, S. M., & Keyes-Elstein, L. (2001). *North Carolina's kindergartners and schools: Summary report.* Chapel Hill, NC: University of North Carolina, Frank Porter Graham Child Development Center. Available: **www.fpg.unc.edu/-SchoolReadiness**

Maxwell, K. L., & Ridley, S. M. (2001, April). *School readiness in North Carolina: Definition, dilemmas, and data.* Paper presented at the Biennial Meeting of the Society for Research in Child Development, Minneapolis, MN.

Meisels, S. J. (1999). Assessing readiness. In R. C. Pianta & M. J. Cox (Eds.), *The transition to kindergarten* (pp. 39–66). Baltimore, MD: Paul H. Brookes Publishing Co., Inc.

Meisels, S. J. (2000). *Developing a system of school readiness assessment.* Paper presented at the National Summit on School Readiness Assessment, Durham, NC.

Muenchow, S. (2001, December). *Implication issues in readiness assessments: Minimizing risks, maximizing benefits.* Paper presented at the Symposium on Assessing the State of State Assessments, Atlanta, GA.

National Education Goals Panel. (1998a). *Principles and recommendations for early childhood assessments.* Washington, DC: National Education Goals Panel.

National Education Goals Panel. (1998b). *Ready schools.* Washington, DC: National Education Goals Panel.

National Research Council and Institute of Medicine. (2000). *From neurons to neighborhoods: The science of early childhood development.* Committee on Integrating the Science of Early Childhood Development. J. P. Schkoff & D. A. Phillips (Eds.), Board on Children, Youth, and Families, Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Phillips, D., & Love, J. (1997). Indicators of school readiness, schooling, and childcare in early to middle childhood. In R. M. Hauser, B. V. Brown, & W. R. Prosser (Eds.), *Indicators of children's well being* (pp.125–151). New York: Russell Sage Foundation.

Ready for School Goal Team. (2000). *School readiness in North Carolina: Strategies for defining, measuring, and promoting success for all children.* Report of the Ready School Goal Team, North Carolina, School Improvement Panel, North Carolina State Board of Education.

Saluja, G., Scott-Little, C., & Clifford, R. M. (2000). Readiness for school: A survey of state policies and definitions. *Early Childhood Research and Practice* [online serial], *2*(2). Available: **http://ecrp.uiuc.edu/v2n2/saluja.html**

Shepard, L. A., Taylor, G. A., & Kagan, S. L. (1996). *Trends in early childhood assessment policies and practices.* Unpublished manuscript.

Zaslow, M., Halle, T., Cabrera, N., Calkins, J., Pitzer, L., & Margie, N. G. (2002). *Child outcome measures in research on childcare quality.* Manuscript submitted for publication.

Zaslow, M., Reidy, M., Moorehouse, M., Halle, T., Calkins, J., & Margie, N. G. (2001, June). *Progress and prospects in the development of indicators of school readiness.* Paper presented at the meeting on Child and Youth Indicators: Accomplishments and Future Directions, National Institutes of Health.

# About the Editors

## Richard M. Clifford

**Senior Scientist, Frank Porter Graham Child Development Institute**
**Research Associate Professor, School of Education,**
**University of North Carolina at Chapel Hill**
**Co-director, National Center for Early Development and Learning and the National Prekindergarten Center**

Richard M. Clifford's training is in Educational Administration with specializations in Political Science and Research. He has taught and served as a principal in public schools. For more than 25 years, he has studied public policies and advised government officials and practitioners on policies affecting children and families. His work focuses on two major areas: public financing of programs for young children and the provision of appropriate learning environments for preschool and early school-age children. Clifford is co-author of a widely used series of instruments for evaluating learning environments for children. In 1993–1994, Clifford helped establish and served as the first director of the Division of Child Development in the North Carolina Department of Human Resources and helped with the design and implementation of the state's Smart Start early childhood initiative. He is a past president of the National Association for the Education of Young Children.

## Sharon Lynn Kagan

**Professor**
**Teachers College, Columbia University**

Sharon Lynn Kagan is the Virginia and Leonard Marx Professor of Early Childhood and Family Policy at Teachers College, Columbia University, and a Professor Adjunct at Yale University's Child Study Center. Author of over 100 articles and 12 books, Kagan's research focuses on the institutions and policies that impact child and family life. Kagan consults with numerous federal and state agencies, congress, governors, and legislators; is a member of 40 national boards; and is Past President of the National Association for the Education of Young Children and Family Support America.

## Catherine Scott-Little

**Expanded Learning Opportunities Project Director**
**The Regional Educational Laboratory at SERVE**

Catherine Scott-Little is the Expanded Learning Opportunities Project Director at the Regional Educational Laboratory at SERVE. Her research has included investigation of wide-scale early childhood assessment systems, child-based outcome standards developed to describe expectations for children prior to entering kindergarten, and evaluations of after-school programs. Scott-Little has worked with a number of states to examine their readiness assessment systems and to develop early learning standards. Prior to joining SERVE, Scott-Little was Deputy Director of a large Head Start program in Texas, and she also served as program coordinator for a child development center serving single-parent, homeless families.

# *About the Authors*

## Tamara Halle

**Senior Research Associate**
**Child Trends**

Tamara Halle is a developmental psychologist and a Senior Research Associate and Content Area Manager for Early Childhood Development at Child Trends. Her research focuses on the interplay among family processes, children's social and cognitive development, and children's school readiness and achievement. Halle has helped to develop survey and observational studies for the Early Childhood Longitudinal Study—Birth Cohort (ECLS-B), a project administered by the National Center for Education Statistics (NCES), and has compiled indicators of child well-being in the area of education and achievement for two editions of *Trends in the Well-Being of America's Children and Youth*. She also directed the production of *Charting Parenthood: A Statistical Portrait of Fathers and Mothers in America*. Along with colleagues at Child Trends, Halle is providing guidance to the states of South Carolina and California on the design of their statewide, community-based early childhood initiatives.

## Gary T. Henry

**Professor**
**Andrew Young School of Policy Studies**
**Georgia State University**

Gary T. Henry is a professor in the Andrew Young School of Policy Studies at Georgia State University. Henry has evaluated a variety of policies and programs, including Pre-K and the HOPE Scholarship in Georgia, as well as school reforms and accountability systems. The author of *Practical Sampling* (1990), *Graphing Data* (1995), and co-author of *Evaluation: An Integrated Framework for Understanding, Guiding, and Improving Policies and Programs* (2000), Henry has also published extensively in the field of evaluation and policy analysis. He received the Evaluation of the Year Award from the American Evaluation Association in 1998 and the Joseph S. Wholey Distinguished Scholarship Award in 2001 from the American Society for Public Administration and the Center for Accountability and Performance.

## John M. Love

**Senior Fellow**
**Mathematica Policy Research**

John Love is a senior fellow at Mathematica Policy Research with 30 years of experience conducting research, program evaluations, and policy studies with early care and education and family programs. He will shortly complete seven years of co-directing the national evaluation of the Early Head Start program for the Administration for Children and Families. Love has extensive experience in measuring child development and well-being throughout the period from infancy to the early elementary grades. He has a special interest in expanding the often narrow conceptualizations of school readiness and has worked extensively with



69

federal, state, and community-level groups on issues of conceptualizing and assessing school readiness. He has served on numerous panels and advisory committees, including the Head Start Performance Measures Technical Work Group, the advisory panel for the evaluation of the Carnegie Corporation's Starting Points initiative, and the DHHS Secretary's Advisory Committee on Head Start Research and Evaluation.

## Susan Muenchow

**Senior Research Scientist**
**American Institutes for Research**

Susan Muenchow is a Senior Research Scientist at American Institutes for Research. Her current work includes directing a project adapting California's Prekindergarten Learning and Development Guidelines to family childcare settings and participating in the evaluation of California's Proposition 10 Initiative and in the National Head Start Impact Study. Her previous work has included serving as the first director of the Florida Partnership for School Readiness and as the executive director of the Florida Children's Forum. She is the co-author, with Dr. Edward Zigler, of *Head Start: America's Most Successful Educational Experiment.*

## Lawrence J. Schweinhart

**Senior Research Scientist**
**High/Scope Educational Research Foundation**

Lawrence J. Schweinhart is an early childhood program researcher and speaker. He has conducted research at the High/Scope Educational Research Foundation in Ypsilanti, Michigan, since 1975 and has chaired its research division since 1989. He currently directs High/Scope's Head Start Quality Research Center, the state evaluation of the Michigan School Readiness Program, and the development and validation of the High/Scope Child Observation Record. He is the lead researcher on the High/Scope Perry Pre-school Study and the High/Scope Pre-school Curriculum Comparison Study. He has served on the Governing Board of the National Association for the Education of Young Children.

## Martha Zaslow

**Vice President for Research**
**Child Trends**

Martha Zaslow is a developmental psychologist and Vice President for Research at Child Trends. Her research takes an ecological perspective, considering the contributions of different contexts to the development of children in low-income families, including the family, childcare, and policy contexts. In studying the role of the family, Zaslow has focused especially on parenting, carrying out observational studies of mother-child interaction in samples of families with a history of welfare receipt. In studying childcare, her work has focused on the use of childcare by families receiving welfare, on nonstandard work hours and childcare among working poor families, and on strategies to improve childcare quality. With respect to the policy context, Zaslow has studied the impacts on children of different welfare reform policies. She is participating in the evaluations of state-level early childhood initiatives aimed at improving children's school readiness in South Carolina and California.

# *About the National Center for Children and Families:*

## Advancing Policy, Education, and Development

The National Center for Children and Families advances the policy, education, and development of children and their families. Housed at Teachers College, Columbia University, the Center produces and applies interdisciplinary research to improve practice and to raise public awareness of social issues that affect the well-being of America's children and families. This work is accomplished through the systematic training of future leaders, scholars, and policy scientists; cutting-edge research and analyses; and dissemination of information to the media, policymakers, and practitioners.

Under the co-direction of Jeanne Brooks-Gunn and Sharon Lynn Kagan, the Center brings together leading scholars from psychology, education, health, family studies, psychiatry, sociology, economics, and political science in the interdisciplinary analyses of complex social phenomena. The Center collaborates with various schools of Columbia University and departments at Teachers College and with centers engaged in similar work nationally and internationally. Its operations are rooted in a commitment to collective engagement in the solution of contemporary social and public problems.

Strategically, the Center does not accept the status quo for children and families. Its existence is predicated on the knowledge that a healthy America depends on socially, emotionally, physically, and intellectually healthy children; productive and loving families; and supportive and empowering communities. Its vision is that these conditions will become reality only through the positive synergy of premier scholarship, relentless public will, and scientifically grounded social strategies. The Center's mission is to evoke these conditions, always mindful that we are not limited by what currently exists, but emboldened by what can and should be.

The work of the Center is accomplished by its faculty and fellows through a set of synergistic activities, including:

- *Research and Publication Opportunities*
- *Training and Fellowship Opportunities*
- *Policy, Legislative, and Dissemination Opportunities*

The Center conducts various kinds of cutting-edge research, ranging from empirical studies to evaluation of intervention and prevention programs to analytic investigation of major issues that affect children and families. This work is conducted with colleagues from research, centers and institutes throughout the nation. In addition, opportunities exist for fellows to work with senior faculty of the National Center for Children and Families on related research and the Center is also affiliated with the Columbia University Institute for Child and Family Policy. Faculty and fellows of the Center regularly present the findings of their research at major forums across the country and publish in distinguished journals and books.



71

With its array of policy briefs, publications, and resource guides, the Center reaches a wide audience, including the media, decision makers, and practitioners. Through testimony to policymakers from the legislative and executive branches of government, active engagement with national research and professional organizations and editorial policy boards, and participation in national study panels, the Center is constantly extending its reach to improve practice and raise public awareness of social issues that affect the well-being of children and families.

To achieve these goals, the Center presently focuses its work on the following five related policy research themes, each of which addresses the prevention of social problems:

- *Early Care and Education*
- *Families*
- *School Transitions and Readiness*
- *Systems/Governance*
- *Neighborhood/Community*

**(Note: Below are brief descriptions of the themes.)**

### *Early Care and Education*

Research has shown that early childhood education has a positive impact on the lives and academic performance of young children. Yet early care and education services are in short supply and only a small percentage of programs provide the kind of high-quality care that produces the best outcomes for children. Work in early care and education concentrates on the analysis of relationships among the supply, quality, and affordability of childcare and early education services in the United States. It also examines the impacts of various prevention programs and interventions, such as Early Head Start and Welfare Reform, on children, families, and the supply and quality of services.

### *Families*

The relationship between work and family life is central to the productivity and well-being of America's citizens. The Center's work in this area focuses on how public policies and programs support adults in their multiple roles. The goal is to understand and influence policies that promote healthy family life, with emphasis placed on fatherhood, unmarried couples, at-risk families, child support enforcement, and family support prevention programs.

### *School Transitions and Readiness*

Preparing children for school and future life success ranks high on the nation's policy agenda. School readiness initiatives and greater attention to children's transitions in early childhood and beyond are emerging as integral parts of education reform and larger social reform movements in support of children and families. The Center's work in this area spans the continuum from transitions in the earliest years of life through middle school, with particular attention given to the intersection of school and family changes and the impact on children's emotional and social well-being.

### *Systems/Governance*

In contrast to other developed nations, America does not have an integrated, comprehensive system of early care and education for its youngest citizens. The Center's work in this area acknowledges the lack of such a system. It examines all the elements, including governance and finance, that need to be addressed in the creation of an early care and education system and includes strategies for outreach and dissemination of information for those attempting to develop a more coherent approach to service delivery for young children and families.

### *Neighborhood/Community*

Healthy children can only develop in the context of healthy neighborhoods and communities. In recognition of this knowledge, most work of the Center includes a neighborhood/community component. The Center's projects focused exclusively in this area examine the influence of neighborhood processes on children of different ages and the effects of residential change on low-income families, with particular attention given to the intersection of neighborhood and family resources and the opportunities and challenges they present for enhancing the well-being of children.



73

# *About the National Center for Early Development and Learning*

The National Center for Early Development and Learning (NCEDL), established in 1996, is the national early childhood research center funded by the U.S. Department of Education. The Center is a component of the Frank Porter Graham Child Development Institute at the University of North Carolina at Chapel Hill. Currently, NCEDL is engaged in a major, multi-state, longitudinal study of (a) the relation between school-related pre-kindergarten experiences and early school outcomes for children and (b) the transition from pre-kindergarten through kindergarten. NCEDL is also involved in a number of other research activities to support the Department and to further research in early childhood education in the U.S.

In recent years, schools have become increasingly involved in providing services for children and families prior to kindergarten entry and the public investment in formal, school-related programs for young children has soared. To date there have been no systematic multi-state studies of the nature and quality of experiences offered to children in these settings or the extent to which variation in experiences relates to child outcomes. Controversy exists about such critical issues as the extent to which academic skills should be taught, the amount and type of teacher training required, the role of families in the programs, and the intensity needed to achieve desired results. Thus NCEDL chose to target its research capability primarily toward addressing this pressing set of issues for our country. The current study aims to fill this gap through a six state, two-year study of 240 pre-kindergarten classrooms and 960 children. Information on quality, practices, and outcomes is being collected through intensive observations, child assessments, interviews, and questionnaires during pre-kindergarten and kindergarten.

In addition to this major study of school-related pre-kindergartens, NCEDL is conducting sub-studies using this same sample, including one that invovles interviews in family's homes and one considering the financing of pre-kindergarten programs. NCEDL also is continuing its data collection efforts around state pre-kindergarten initiatives and other research requested by the U.S. Department of Education. Dissemination of relevant findings and policy recommendations is a high priority for NCEDL.

**BEST COPY AVAILABLE**

# About SERVE

SERVE, directed by Dr. John R. Sanders, is an education organization with the mission to promote and support the continuous improvement of educational opportunities for all learners in the Southeast. The organization's commitment to continuous improvement is manifest in an applied research-to-practice model that drives all of its work. Building on theory and craft knowledge, SERVE staff members develop tools and processes designed to assist practitioners and policymakers with their work, ultimately, to raise the level of student achievement in the region. Evaluation of the impact of these activities combined with input from affected stakeholders expands SERVE's knowledge base and informs future research.

This vigorous and practical approach to research and development is supported by an experienced staff strategically located throughout the region. This staff is highly skilled in providing needs assessment services, conducting applied research in schools, and developing processes, products, and programs that inform educators and increase student achievement. In the last three years, in addition to its basic research and development work with over 170 southeastern schools, SERVE staff provided technical assistance and training to more than 18,000 teachers and administrators across the region.

SERVE is governed by a board of directors that includes the governors, chief state school officers, educators, legislators, and private sector leaders from Alabama, Florida, Georgia, Mississippi, North Carolina, and South Carolina.

At the core of SERVE's business is the operation of the Regional Educational Laboratory. Funded by the U.S. Department of Education's National Institute for Education Sciences, the Regional Educational Laboratory for the Southeast is one of ten programs providing research-based information and services to all 50 states and territories. These Laboratories form a nationwide education knowledge network, building a bank of information and resources shared nationally and disseminated regionally to improve student achievement locally. SERVE's National Leadership Area, Expanded Learning Opportunities, focuses on improving student outcomes through the use of exemplary pre-K and extended-day programs.

In addition to the Lab, SERVE operates the Southeast Eisenhower Regional Consortium for Mathematics and Science Education and the SouthEast Initiatives Regional Technology in Education Consortium (SEIR◆TEC). SERVE also administers a subcontract for the Region IV Comprehensive Center and has additional funding from the Department to provide services in migrant education and to operate the National Center for Homeless Education and the Adjunct ERIC Clearinghouse on Homeless Education.

Together, these various elements of SERVE's portfolio provide resources, services, and products for responding to regional and national needs. Program areas include:

- Assessment, Accountability, and Standards
- Children, Families, and Communities
- Education Leadership
- Education Policy

- Improvement of Science and Mathematics Education
- Reading and School Improvement
- Technology in Learning

In addition to the program areas, the SERVE Evaluation Unit supports the evaluation activities of the major grants and contracts and provides contracted evaluation services to state and local education agencies in the region. The Technology Support Group provides SERVE staff and their constituents with IT support, technical assistance, and software applications. Through its Publications Unit, SERVE publishes a variety of studies, training materials, policy briefs, and program products. Among the many products developed at SERVE, two receiving national recognition include *Achieving Your Vision of Professional Development*, honored by the National Staff Development Council, and *Study Guide for Classroom Assessment: Linking Instruction and Assessment*, honored by Division H of AERA. Through its programmatic, technology, evaluation, and publishing activities, SERVE provides contracted staff development and technical assistance in specialized areas to assist education agencies in achieving their school improvement goals.

SERVE's main office is at the University of North Carolina at Greensboro, with major staff groups located in Tallahassee, Florida, and Atlanta, Georgia, as well as satellite offices in Durham, North Carolina, and Shelby, Mississippi. Unique among the ten Regional Educational Laboratories, SERVE employs a full-time policy analyst to assist the chief state school officer at the state education agencies in each of the states in the SERVE region. These analysts act as SERVE's primary liaisons to the state departments of education, providing research-based policy services to state-level education.

***SERVE Main Office • P.O. Box 5367 • Greensboro, NC 27435***
***800-755-3277 • 336-315-7400 • Fax 336-315-7457***
***John R. Sanders, Ed.D. • Executive Director***
***www.serve.org***

# *Assessing the State of State Assessments:*

## Perspectives on Assessing Young Children

Wide-scale assessment systems to collect data on the characteristics of large numbers of young children have become increasingly common in recent years, and educators, both at the federal and the state level, have seen increasing pressure to assess children at younger ages. With this increasing pressure has come concern regarding the purpose of assessments, the nature of assessment processes, and the implications for how the data are being used. In response, policymakers, practitioners, and researchers have struggled to design and implement early childhood assessment systems that are valid, reliable, fair, and practical.

*Assessing the State of State Assessments: Perspectives on Assessing Young Children* is a compilation of perspectives on assessment issues written by authors deeply involved in developing and implementing wide-scale assessment systems. Taken together, the chapters provide practical and theoretical insights into four areas critical in developing such systems: design, instrumentation, implementation, and data utilization. In the chapters, noted experts describe not only the challenges inherent in early childhood assessment but also strategies for meeting those challenges.



SRASSA

SERVE™
*Improving Learning through Research & Development*

**U.S. Department of Education**
*Office of Educational Research and Improvement (OERI)*
*National Library of Education (NLE)*
*Educational Resources Information Center (ERIC)*

# NOTICE

## Reproduction Basis

| | |
|---|---|
| | This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form. |
| X | This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket"). |


P2031310